PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/31, 15/54, 9/10, 1/21, C07K 14/315, C12P 19/04, A61K 39/09, C12Q 1/14, 1/68, G01N 33/569

A2

(11) International Publication Number: **WO 00/05378**

(43) International Publication Date: 3 February 2000 (03.02.00)

(21) International Application Number: PCT/NL99/00460

(22) International Filing Date: 19 July 1999 (19.07.99)

(30) Priority Data:

98202465.5	22 July 1998 (22.07.98)	EP
98202467.1	22 July 1998 (22.07.98)	EP

(71) Applicant *(for all designated States except US)*: STICHTING DIENST LANDBOUWKUNDIG ONDERZOEK [NL/NL]; Bornsesteeg 53, NL–6708 PD Wageningen (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: SMITH, Hilda, Elizabeth [NL/NL]; Golfpark 98, NL–8241 AG Lelystad (NL).

(74) Agent: OTTEVANGERS, S., U.; Vereenigde Octrooibureaux, Nieuwe Parklaan 97, NL–2587 BN The Hague (NL).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: *STREPTOCOCCUS SUIS* VACCINES AND DIAGNOSTIC TESTS

(57) Abstract

The invention relates to *Streptococcus suis* infections of pigs, to vaccines directed against those infections and to tests for diagnosing *Streptococcus suis* infections. The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derivated thereof. The invention furthermore provides a nucleic acis probe or primer allowing species or serotype specific detection of *Streptococcus suis*. The invention also provides a *Streptococcus suis* antigen and vaccine derived thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Title: *Streptococcus suis* vaccines and diagnostic tests.

The invention relates to *Streptococcus* infections of pigs, to vaccines directed against those infections, to tests for diagnosing *Streptococcus* infections and to the field of bacterial vaccines, more in particular to vaccines directed against *Streptococcus* infections.

Streptococcus species, of which there are a large variety causing infections in domestic animals and man, are often grouped according to Lancefield's groups. Typing according to Lancefield occurs on the basis of serological determinants or antigens that are among others present in the capsule of the bacterium and allows for only an approximate determination, often bacteria from a different group show cross-reactivity with each other, while other Streptococci can not be assigned a group-determinant at all. Within groups, further differentiation is often possible on the basis of serotyping; these serotypes further contribute to the large antigenic variability of Streptococci, a fact that creates an array of difficulties within diagnosis of and vaccination against Streptococcal infections.

Lancefield group A *Streptococcus species* (GAS, *Streptococcus pyogenes*), are common with children, causing nasopharyngeal infections and complications thereof. Among animals, especially cattle are susceptible to GAS, whereby often mastitis is found.

Group A streptococci are the etiologic agents of streptococcal pharyngitis and impetigo, two of the commonest bacterial infections in children, as well as a variety of less common but potentially life-threatening infections, including soft tissue infections, bacteraemia, and pneumonia. In addition, GAS are uniquely associated with the postinfectious autoimmune syndromes of acute rheumatic fever and poststreptococcal glomerulonephritis.

Several recent reports suggest that the incidence both of serious infections due to GAS and of acute rheumatic fever has

increased during the past decade, focusing renewed interest on defining the attributes or virulence factors of the organism that may play a role in the pathogenesis of these diseases. GAS produce several surface components and extracellular products that may be important in virulence. The major surface protein, M protein, has been studied in the most detail and has been shown convincingly to play a role in both virulence and immunity. Isolates rich in M protein are able to grow in human blood, a property thought to reflect the capacity of M protein to interfere with phagocytosis, and these isolates tend to be virulent in experimental animals.

Lancefield group B *Streptococcus* (GBS) are most often seen with cattle, causing mastitis, however, human infants are susceptible as well, often with fatal consequences. Group B streptococci (GBS) constitute a major cause of bacterial sepsis and meningitis among human neonates born in the United States and Western Europe and are emerging as significant neonatal pathogens in developing countries as well.

It is estimated that GBS strains are responsible for 10,000 to 15,000 cases of invasive infection in neonates in the United States alone. Despite advances in early diagnosis and treatment, neonatal sepsis due to GBS continues to carry a mortality rate of 15 to 20%. In addition, survivors of GBS meningitis have 30 to 50% incidence of long-term neurologic sequelae. The increasing recognition over the past two decades of GBS as an important pathogen for human infants has generated renewed interest in defining the bacterial and host factors important in virulence of GBS and in the immune response to GBS infection.

Particular attention has focused on the capsular polysaccharide as the predominant surface antigen of the organisms. In a modification of the system originally developed by Rebecca Lancefield, GBS strains are serotyped on the basis of antigen differences in their capsular polysaccharides and the presence or absence of serologically defined C proteins. While GBS isolated from non-human sources

often lack a serologically detectable capsule, a large majority of strains associated with neonatal infection belong to one of four major capsular serotypes, 1a, 1b, II or III. The capsular polysaccharide forms the outermost layer around the exterior of the bacterial cell, superficial to the cell wall. The capsule is distinct from the cell wall-associated group B carbohydrate. It has been suggested that the presence of sialic acid in the capsule of bacteria that cause meningitis is important for these bacteria to breach the blood-brain barrier. Indeed, in S. agalactiae sialic acid has shown to be critical for the virulence function of the type III capsule. The capsule of S. suis serotype is composed of glucose, galactose, N-acetylglucosamine, rhamnose and sialic acid.

The group B polysaccharide, in contrast to the type-specific capsule, is present on all GBS strains and is the basis for serogrouping of the organisms into Lancefield's group B. Early studies by Lancefield and co-workers showed that antibodies raised in rabbits against whole GBS organisms protected mice against challenge with strains of homologous capsular type, demonstrating the central role of the capsular polysaccharide as a protective antigen. Studies in the 1970s by Baker and Kasper demonstrated that cord blood of human infants with type III GBS sepsis uniformly had low or undetectable levels of antibodies directed against the type III capsule, suggesting that a deficiency of anticapsular antibody was a key factor in susceptibility of human neonates to GBS disease.

Lancefield group C infections, such as those with *S. equi, S. zooepidemicus, S. dysgalactiae*, and others are mainly seen with horse, cattle and pigs, but can also cross the species barrier to humans. Lancefield group D (*S. bovis*) infections are found with all mammals and some birds, sometimes resulting in endocarditis or septicaemia.

Lancefield groups E, G, L, P, U and V (*S. porcinus, S, canis, S. dysgalactiae*) are found with various hosts, causing

neonatal infections, nasopharyngeal infections or mastitis.

Within Lancefield groups R, S, and T, (and with ungrouped types) *S. suis* is found, an important cause of meningitis, septicemia, arthritis and sudden death in young pigs. Incidentally, it can also cause meningitis in man.

Streptococcus suis is an important cause of meningitis, septicemia, arthritis and sudden death in young pigs (4, 46). Incidentally, it can also cause meningitis in man (1). *S.suis* strains are usually identified and classified by their morphological, biochemical and serological characteristics (58, 59, 46). Serological classification is based on the presence of specific antigenic polysaccharides. So far, 35 different serotypes have been described (9, 56, 14). In several European countries, *S. suis* serotype 2 is the most prevalent type isolated from diseased pigs, followed by serotypes 9 and 1. Serological typing of *S. suis* is carried out using different types of agglutination tests. In these tests, isolated and biochemically characterised *S. suis* cells are agglutinated with a panel of 35 specific sera. These methods are very laborious and time-consuming.

Little is known about the pathogenesis of the disease caused by *S. suis,* let alone about its various serotypes such as type 2. Various bacterial components, such as extracellular and cell-membrane associated proteins, fimbriae, haemaglutinins, and haemolysin have been suggested as virulence factors (9, 10, 11, 15, 16, 47, 49). However, the precise role of these protein components in the pathogenesis of the disease remains unclear (37). It is well known that the polysaccharidic capsule of various Streptococci and other gram-positive bacteria plays an important role in pathogenesis (3, 6, 35, 51, 52). The capsule enables these micro-organisms to resist phagocytosis and is therefore regarded as an important virulence factor. Recently, a role of the capsule of *S. suis* in the pathogenesis was suggested as well (5). However, the structure, organisation and functioning of the genes responsible for capsule polysaccharide synthesis (*cps*) in *S. suis* is unknown. Within *S. suis* serotypes

1 and 2 strains can differ in virulence for pigs (41, 45, 49). Some type 1 and 2 strains are virulent, other strains are not. Because both virulent and non-virulent strains of serotype 1 and 2 strains are fully encapsulated, it may even be that capsule is not a relevant factor required for virulence.

Attempts to control *S. suis* infections or disease are still hampered by the lack of knowledge about the epidemiology of the disease and the lack of effective vaccines and sensitive diagnostics. It is well known and generally accepted that the polysaccharidic capsule of various Streptococci and other gram-positive bacteria plays an important role in pathogenesis. The capsule enables these micro-organisms to resist phagocytosis and is therefore regarded as an important virulence factor.

Compared to encapsulated *S. suis* strains, non-encapsulated *S. suis* strains are phagocytosed by murine polymorphonuclear leucocytes to a greater degree. Moreover, an increase in thickness of capsule was noted for *in vivo* grown virulent strains while no increase was observed for avirulent strains. Therefor, these data again demonstrate the role of the capsule in the pathogenesis for *S. suis* as well.

Ungrouped *Streptoccus species*, such as *S. mutans*, causing carries with humans, *S, uberis*, causing mastitis with cattle, and *S. pneumonia*, causing major infections in humans, and *Enterococcus faecilalis* and *E. faecium*, further contributed to the large group of Streptococci.

Streptococcus pneumoniae (the pneumococcus) is a human pathogen causing invasive diseases, such as pneumonia, bacteraemia, and meningitis. Despite the availability of antibiotics, pneumococcal infections remain common and can still be fatal, especially in high-risk groups, such as young children and elderly people. Particularly in developing countries, many children under the age of five years die each year from pneumococcal pneumonia. *S. pneumoniae* is also the leading cause of otitis media and sinusitis. These infections are less serious, but nevertheless incur substantial medical

costs, especially when leading to complications, such as permanent deafness. The normal ecological niche of the pneumococcus is the nasopharynx of man. The entire human population is colonised by the pneumococcus at one time or another, and at a given time, up to 60% of individuals may be carriers. Nasopharyngeal carriage of pneumococci by man is often accompanied by the development of protection to infection by the same serotype. Most infections do not occur after prolonged carriage but follow the acquisition of recently acquired strains. Many bacteria contain surface polysaccharides which act as a protective layer against the environment. Surface polysaccharides of pathogenic bacteria usually make the bacteria resistant to the defense mechanisms of the host, e.g., the lytic action of serum or phagocytosis. In this respect, the serotype-specific capsular polysaccharide (CP) of *Streptococcus pneumoniae*, is an important virulence factor. Unencapsulated strains are avirulent, and antibodies directed against the CP are protective. Protection is serotype specific; each serotype has its own, specific CP structure. Ninety different capsular serotypes have been identified. Currently, CPs of 23 serotypes are included in a vaccine.

Vaccines directed against *Streptococcus* infections in general aim at utilising an immune response directed against the polysaccharide capsule of the various *Streptococcus species*, especially since the capsule is considered a main virulence factor for these bacteria. The capsule, during infection, provides resistance to phagocytosis and thus promotes the escape of the bacteria from the immune system of the host, protecting the bacteria by elimination by macrophages and neutrophils.

The capsule particularly confers the bacterium resistance to complement-mediated opsonophagocytosis. In addition, some bacteria express capsular polysaccharides (CPs) that mimic host molecules, thereby avoiding the immune system of the host. Also, even when the bacteria have been phagocytosed, intracellular killing is hampered by the presence of a

capsule.

It is in general thought that only when the host has antibodies or other serum-factors directed against capsule antigens, the bacterium will get recognised by the immune system through the anticapsular-antibodies or serum-factors bound to its capsule, and will, through opsonisation, get phagocytosed and killed.

However, these antibodies are serotype-specific, and will often only confer protection against only one of the many serotypes known within a group of *Streptococci*.

For example, current commercially available *S. suis* vaccines, which are in general based on whole-cell-bacterial preparations, or on capsule-enriched fractions of *S. suis*, confer only limited protection against heterologous strains. Also, the current pneumococcal vaccine, licensed in the United States in 1983, consists of purified CPs of 23 pneumococcal serotypes whereas at least 90 CP types exist.

The composition of this pneumococcal vaccine was based on the frequency of the occurrence of disease isolates in the US and cross-reactivity between various serotypes. Although this vaccine protects healthy adults against infections caused by serotypes included in the vaccine, it fails to raise a protective immune response in infants younger than 18 months and it is less effective in elderly people. In addition, the vaccine confers only limited protection in patients with immunodeficiencies and haematology malignancies.

In the light of above, improved vaccines are needed against *Streptococcus* infections. Much attention is being paid at producing CP vaccines by producing the relevant polysaccharides via chemical or recombinant means. However, chemical synthesis of polysaccharides is costly, and capsular polysaccharide synthesis by recombinant means necessitates knowledge about the relevant genes, which are not always available and need to de determined for each and every relevant serotype.

The invention provides an isolated or recombinant nucleic acid encoding a capsular (*cps*) gene cluster of *Streptococcus suis*. Biosynthesis of capsule polysaccharides in general has been studied in a number of Gram-positive and Gram-negative bacteria (32). In Gram-negative bacteria, but also in a number of gram-positive bacteria, genes which are involved in the biosynthesis of polysaccharides are clustered at a single locus. *Streptococcus suis* capsular genes as provided by the invention show a common genetic organisation involving three distinct regions. The central region is serotype specific and encodes enzymes responsible for the synthesis and polymerisation of the polysaccharides. This region is flanked by two regions conserved in *Streptococcus suis* which encode proteins for common functions such as transport of the polysaccharide across the cellular membrane. However, in between species, only low homologies exist, hampering easy comparison and detection of seemingly similar genes. Knowing the nucleic acid encoding the flanking regions allows type-specific determination of nucleic acid of the central region of *Streptococcus suis* serotypes, as for example described in the experimental part of the description of the invention.

The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derived thereof. Such a nucleic acid is for example provided by hybridising chromosomal DNA derived from any one of the *Streptococcus suis* serotypes to a nucleic acid encoding a gene derived from a *Streptococcus suis* serotype 1, 2 or 9 capsular gene cluster, as provided by the invention (see for example Tables 4 and 5) and cloning of (type-specific) genes as for example described in the experimental part of the description. At least 14 open reading frames are identified. Most of the genes belong to a single transcriptional unit, identifying a co-ordinate control of these genes, they, and the enzymes and proteins they encode, act in concert to provide the capsule with the relevant polysaccharides. The invention provides *cps* genes and proteins

encoded thereof involved in regulation (CpsA), chain length determination (CpsB, C), export (CpsC) and biosynthesis (CpsE, F, G, H, J, K). Although the overall organisation seemed at first glance to be similar to that of the *cps* and *eps* gene clusters of a number of Gram-positive bacteria (19, 32, 42), overall homologies are low (see table 3). The region involved in biosynthesis is located at the centre of the gene cluster and is flanked by two regions containing genes with more common functions.

The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 2 or a gene or gene fragment derived thereof, preferably as identified in Figure 3. Genes in this gene cluster are involved in polysaccharide biosynthesis of capsular components and antigens. For a further description of such genes see for example Table 2 of the description, for example a cpsA gene is provided functionally encoding regulation of capsular polysaccharide synthesis, whereas cpsB and cpsC are functionally involved in chain in chain length determination. Other genes, such as cpsD, E, F, G, H, I, J, K and related genes, are involved in polysaccharide syntheses, functioning for example as glucosyl- or glycosyltransferase. The cpsF, G, H, I, J genes encode more type-specific proteins than the flanking genes which are found more-or-less conserved throughout the species and can serve as base for selection of primers or probes in PCR-amplification or cross-hybridisation experiments for subsequent cloning.

For example, the invention further provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 1 or a gene or gene fragment derived thereof, preferably as identified in Figure 4.

In addition, the invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 9 or a gene or gene fragment derived thereof, preferably as identified in Figure 5.

Furthermore, the invention provides for example a fragment or parts thereof of the *cps* locus, involved in the capsular polysaccharide biosynthesis, of *S. suis,* exemplified in the experimental part for serotype 1, 2 or 9, and allows easy identification or detection of related fragments derived of other serotype of *S. suis*.

The invention provides a nucleic acid probe or primer derived from a nucleic acid according to the invention allowing species or serotype specific detection of *Streptococcus suis*. Such a probe or primer (herein used interchangeably) is for example a DNA, RNA or PNA (peptide nucleic acid) probe hybridising with capsular nucleic acid as provided by the invention. Species specific detection is provided preferably by selecting a probe or primer sequence from a species-specific region (e.g. flanking region) whereas serotype specific detection is provided preferably by selecting a probe or primer sequence from a type-specific region (e.g. central region) of a capsular gene cluster as provided by the invention. Such a probe or primer can be used in a further unmodified form, for example in cross-hybridisation or polymerase-chain reaction (PCR) experiments as for example described in the experimental part of the description of the invention. Herein the invention provides the isolation and molecular characterisation of additional type specific *cps* genes of *S. suis* types 1 and 9. In addition, we describe the genetic diversity of the *cps* loci of serotypes 1, 2 and 9 among the 35 *S. suis* serotypes yet known. Type-specific probes are identified. Also, a type-specific PCR for for example serotype 9 is provided, being a rapid, reliable and sensitive assay, which is used directly on nasal or tonsillar swabs or other samples of infected or carrier animals.

The invention also provides a probe or primer according to the invention further provided with at least one reporter molecule. Examples of reporter molecules are manifold and known in the art, for example a reporter molecule can comprise

additional nucleic acid provided with a specific sequence (e.g. oligo-dT) hybridising to a corresponding sequence to which hybridisation can easily be detected for example because it has been immobilised to a solid support.

Yet other reporter molecules comprise chromophores, e.g. fluorochromes for visual detection, for example by light microscopy or fluorescent in situ hybridisation (FISH) techniques, or comprise an enzyme such as horseradish peroxidase for enzymatic detection, e.g in enzyme-linked assays (EIA). Yet other reporter molecules comprise radioactive compounds for detection in radiation-based-assays.

In a preferred embodiment of the invention, at least one probe or primer according to the invention is provided (labelled) with a reporter molecule and a quencher molecule, providing together with unlabeled probe or primer a PCR-based test allowing rapid detection of specific hybridisation.

The invention further provides a diagnostic test or test kit comprising a probe or primer as provided by the invention. Such a test or test kit, for example a cross-hybridisation test or PCR-based test, is advantageously used in rapid detection and/or serotyping of *Streptococcus suis.*

The invention furthermore provides a protein or fragment thereof encoded by a nucleic acid according to the invention. Examples of such a protein or fragment are for example proteins described in for example Table 2 of the description, for example a cpsA protein is provided functionally encoding regulation of capsular polysaccharide synthesis, whereas cpsB and cpsC are functionally involved in chain in chain length determination. Other proteins or functional fragments thereof as provided by the invention, such as cpsD, E, F, G, H, I, J, K and related proteins, are involved in polysaccharide biosynthesis, functioning for example as glucosyl- or glycosyltransferase in polysaccharide biosynthesis of *Streptococcus suis* capsular antigen.

The invention furthermore provides a method to produce a *Streptococcus suis* capsular antigen comprising using a protein

or functional fragment thereof as provided by the invention, and provides therewith a *Streptococcus suis* capsular antigen obtainable by such a method. A comparison of the predicted amino acid sequences of the *cps2* genes with sequences found in the databases allowed the assignment of functions to the open reading frames. The central region contains the type specific glycosyltransferases and the putative polysaccharide polymerase. This region is flanked by two regions encoding for proteins with common functions, such as regulation and transport of polysaccharide across the membrane. Biosynthesis of Streptococcus capsular polysaccharide antigen using a protein or functional fragment thereof is advantageously used in chemo-enzymatic synthesis and the development of vaccines which offer protection against serotype-specific Streptococcal disease, and is also advantageously used in the synthesis and development of multivalent vaccines against Streptococcal infections. Such vaccines elicit anticapsular antibodies which confer protection.

Furthermore, the invention provides an acapsular *Streptococcus* mutant for use in a vaccine, a vaccine strain derived thereof and a vaccine derived thereof. Surprisingly, and against the grain of common doctrine, the invention provides use of a *Streptococcus* mutant deficient in capsular expression in a vaccine.

Acapsular *Streptococcus* mutants have long been known in the art and can be found in nature. Griffith (J. Hyg. 27:113-159, 1928) demonstrated that pneumococci could be transformed from one type to another. If he injected live rough (acapsular or unencapsulated) type 2 pneumococci into mice, the mice would survive. If, however, he injected the same dose of live rough type 2 mixed with heat-killed smooth (encapsulated) type 1 into a mouse, the mouse would die, and from the blood he could isolate live smooth type 1 pneumococci. At that time, the significance of this transforming principle was not understood. However, understanding came when it was shown that

DNA constituted the genetic material responsible for phenotypic changes during transformation.

Streptococcus mutants deficient in capsular expression are found in several forms. Some are fully deficient and have no capsule at all, others form a deficient capsule, characterised by a mutation in a capsular gene cluster. Deficiency can for instance include capsular formation wherein the organization of the capsular material has been re-arranged, as for example demnosrable by electron microscopy. Yet others have a nearly fully developed capsule which is only deficient in a particular sugar component.

Now, after much advance of biotechnology and despite the fact that little is still known about the exact localisation and sequence of genes involved in capsular synthesis in Streptococci, it is possible to create mutants of Streptococci, for example by homologous recombination or transposon mutagenesis, which has for example been done for GAS (Wessels et al., PNAS 88:8317-8321, 1991), for GBS (Wesels et al., PNAS 86: 8983-8987, 1989), for *S. suis* (Smith, ID-DLO Annual report 1996, page 18-19; Charland et al., Microbiol. 144:325-332, 1998) and for *S. pneumonia* (Kolkman et al., J. Bact. 178:3736-3741, 1996). Such recombinant derived mutants, or isogenic mutants, can easily be compared with the wild-type strains from which they have been derived.

In a preferred embodiment, the invention provides use of a recombinant-derived *Streptococcus* mutant deficient in capsular expression in a vaccine. Recombinant techniques useful in producing such mutants are for example homologous recombination, transposon mutagenises, and others, whereby deletions, insertions or (point)-mutations are introduced in the genome. Advantages of using recombinant techniques are the stability of the obtained mutants (especially with homologous recombination and double cross-over techniques), and the knowledge about the exact site of the deletion, mutation or insertion.

In a much preferred embodiment, the invention provides a

stable mutant deficient in capsular expression obtainable for example through homologous recombination or cross over integration events. Examples of such a mutant can be found in the experimental part of this description, for example mutant 10cpsB or 10cpsEF is such a stable mutant as provided by the invention.

The invention also provides a *Streptococcus* vaccine strain and vaccine that has been derived from a *Streptococcus* mutant deficient in capsular expression. In general, said strain or vaccine is applicable within the whole range of Streptococcal infections, be it for those with animals or man or with zoonotic infections. It is of course now possible to first select a common vaccine strain and derive a *Streptococcus* mutant deficient in capsular expression thereof for the selection of a vaccine strain and use in a vaccine according to the invention.

In a preferred embodiment, the invention provides use of a *Streptococcus* mutant deficient in capsular expression in a vaccine wherein said *Streptococcus* mutant is selected from the group composed of *Streptococcus* group A, *Streptococcus* group B, *Streptococcus suis* and *Streptococcus pneumonia*. Herewith the invention provides vaccine strains and vaccines for use with these notoriously heterologous Streptococci, of which a multitude of serotypes exist. With a vaccine as provided by the invention that is derived from a specific *Streptococcus* mutant that deficient in capsular expression, the difficulties relating to lack of heterologous protection can be circumvented since these mutants do nor rely on capsular antigens per se to induce protection.

In a preferred embodiment, said vaccine strain is selected for its ability to survive or even replicate in an immune-competent host or host cells and thus can persist for a certain period, varying from 1-2 days to more than one or two weeks, in a host, despite its deficient character.

Although an immunodeficient host will support replication of a wide range of bacteria that are deficient in one or more

virulence factors, in general it is considered a characteristic of pathogenicity of Streptococci that they can survive for certain periods or replicate in a normal host or host cells such as macrophages. For example, Wiliams and Blakemore (Neuropath. Appl. Neurobiol.: 16, 345-356, 1990; Neuropath. Appl. Neurobiol.: 16, 377-392, 1990; J. Infect. Dis.: 162, 474-481, 1990) show that both polymorphonuclear cells and macrophage cells are capable of phagocytosing pathogenic *S. suis* in pigs lacking anti-*S. suis* antibodies, only pathogenic bacteria could survive and multiply inside macrophages and the pig.

In a preferred embodiment, the invention, however, provides a deficient or avirulent mutant or vaccine strain which is capable of surviving at least 4-5 days, preferably at least 8-10 days in said host, thereby allowing the development of a solid immune response to subsequent *Streptococcus* infection,

Due to its persistent but avirulent character, a *Streptococcus* mutant or vaccine strain as provided by the invention is well suited to generate specific and/or long-lasting immune responses against Streptococcal antigens, moreover because possible specific immune responses of the host directed against a capsule are relatively irrelevant because a vaccine strain as provided by the invention is in general not recognised by such antibodies.

In addition, the invention provides a *Streptococcus* vaccine strain according the invention which strain comprises a mutant capable of expressing a *Streptococcus* virulence factor or antigenic determinant.

In a preferred embodiment, the invention provides a *Streptococcus* vaccine strain according to the invention which strain comprises a mutant capable of expressing a *Streptococcus* virulence factor wherein said virulence factor or antigenic determinant is selected from a group of cellular components, such as muramidase-released protein (MRP) extracellular factor (EF) and cell-membrane associated

proteins, 60kDA heat shock protein, pneumococcal surface protein A (Psp A), pneumolysin, C protein, protein M, fimbriae, haemagglutinins and haemolysin or components functionally related thereto.

In a much preferred embodiment, the invention provides a *Streptococcus* vaccine strain according to the invention which strain comprises a mutant capable of over-expressing said virulence factor. In this way, the invention provides a vaccine strain for incorporation in a vaccine which specifically causes a host to provide a immune response directed against antigenically important determinants of virulence (listed above), thereby providing specific protection directed against said determinants. Over-expression can for example be achieved by cloning the gene involved behind a strong promoter, which is for example constitutionally expressed in a multicopy system, either in a plsamid or via intergration in a genome.

In yet another embodiment, the invention provides a *Streptococcus* vaccine strain according to the invention which comprises a mutant capable of expressing a non-*Streptococcus* protein. Such a vector-*Streptococcus* vaccine strain allows, when used in a vaccine, protection against other pathogens than Streptococcus.

Due to its persistent but avirulent character, a *Streptococcus* vaccine strain or mutant as provided by the invention is well suited to generate specific and long-lasting immune responses, not only against Streptococcal antigens, but also against other antigens when these are expressed by said strain. Especially antigens derived from another pathogen are now expressed without the detrimental effects of said antigen or pathogen which would otherwise have harmed the host.

An example of such a vector is a *Streptococcus* vaccine strain or mutant wherein said antigen is derived from a pathogen, such as *Actinobacillus pleuropneumonia, Mycoplasmatae, Bordetella, Pasteurella, E. coli, Salmonella, Campylobacter, Serpulina* and others.

The invention also provides a vaccine comprising a *Streptococcus* vaccine strain or mutant according to the invention and further comprising a pharmaceutically acceptable carrier or adjuvant. Carriers or adjuvants are well known in the art, examples are phosphate buffered saline, physiological salt solutions, (double-)oil-in-water-emulsions, aluminumhydroxide, Specol, block- or co-polymers, and others.

A vaccine according to the invention can comprise a vaccine strain either in a killed or live form. For example, a killed vaccine comprising a strain having (over)expressed a Streptococcal or heterologous antigen or virulence factor is very well suited for eliciting an immune response. In a preferred embodiment, the invention provides a vaccine wherein said strain is live, due to its persistent but avirulent character, a *Streptococcus* vaccine strain as provided by the invention is well suited to generate specific and long-lasting immune responses.

Now that a Streptococcal vaccine is provided by the invention, the invention also provides a method for controlling or eradicating a Streptococcal disease in a population comprising vaccinating subjects in said population with a vaccine according to the invention.

In a preferred embodiment, a method for controlling or eradicating a Streptococcal disease is provided comprising testing a sample, such as a blood sample, or nasal or throat swab, faeces, urine, or other samples such as can be sampled at or after slaughter, collected from at least one subject, such as an infant or a pig, in a population partly or wholy vaccinated with a vaccine according to the invention for the presence of encapsulated Streptococcal strains or mutants. Since a vaccine strain or mutant according to the invention is not pathogenic, and can be distinguished from wild-type strains by capsular expression, the detection of (fully) encapsulated Streptococcal strains indicates that wild-type infections are still present. Such wild-type infected subjects can than be isolated from the remainder of the population

until the infection has passed away. With domestic animals, such as pigs, it is even possible to remove the infected subject from the population as a whole by culling. Detection of wild-type strains can be achieved via traditional culturing techniques, or by rapid detection techniques such as PCR detection.

In yet another embodiment, the invention provides a method for controlling or eradicating a Streptococcal disease comprising testing a sample collected from at least one subject in a population partly or wholly vaccinated with a vaccine according to the invention for the presence of capsule-specific antibodies directed against Streptococcal strains. Capsule specific antibodies can be detected with classical techniques known in the art, such as used for Lancefield's group typing or serotyping.

A much preferred embodiment of a method provided by the invention for controlling or eradicating a Streptococcal disease in a population comprises vaccinating subjects in said population with a vaccine according to the invention and testing a sample collected from at least one subject in said population for the presence of encapsulated Streptococcal strains and/or for the presence of capsule-specific antibodies directed against Streptococcal strains.

For example, a method is provided according to the invention wherein said Streptococcal disease is caused by *Streptococcus suis*.

The invention also provides a diagnostic assay for testing a sample for use in a method according to the invention comprising at least one means for the detection of encapsulated Streptococcal strains and/or for the detection of capsule-specific antibodies directed against Streptococcal strains.

The invention furthermore provides a vaccine comprising an antigen according to the invention and further comprising a suitable carrier or adjuvant. The immunogenicity of a capsular antigen provided by the invention is for example increased by

linking to a carrier (such as a carrier protein), allowing the recruitment of T-cell help in developing an immune response.

The invention further provides a recombinant micro-organism provided with at least a part of a capsular gene cluster derived from *Streptococcus suis*. The invention provides for example a lactic acid bacterium provided with at least a part of a capsular gene cluster derived from *Streptococcus suis*. Various food-grade lactic acid bacteria (Lactococcus lactis, Lactobacillus casei, Lactobacillus plantarium and Streptococcus gordonii) have been used as delivery systems for mucosal immunization. It has now been shown that oral (or mucosal) administration of recombinant L. lactis, Lactobacillus, and Streptococcus gordonii can elicit local IgA and /or IgG antibody responses to an expressed antigen. The use of oral routes for immunization against infective diseases is desirable because oral vaccines are easier to administer, have higher compliance rates, and because mucosal surfaces are the portals of entry for many pathogenic microbial agents. It is within the skill of the artisan to provide such micro-organisms with (additional) genes.

The invention further provides a recombinant *Streptococcus suis* mutant provided with a modified capsular gene cluster. It is within the skill of the artisan to swap genes within a species. In a preferred embodiment, an avirulent *Streptococcus suis* mutant is selected to be provided with at least a part of a modified capsular gene cluster according to the invention.

The invention further provides a vaccine comprising a micro-organism or a mutant provided by the invention. An advantage of such a vaccine over currently used vaccines is that they comprise accurately defined micro-organisms and well-characterised antigens, allowing accurate determination of immune responses against various antigens of choice.

The invention is further explained in the experimental part of this description without limiting the invention thereto.

Experimental part

MATERIAL AND METHODS

Bacterial strains and growth conditions.
The bacterial strains and plasmids used in this study are listed in Table 1. *S. suis* strains were grown in Todd-Hewitt broth (code CM189, Oxoid), and plated on Columbia agar blood base (code CM331, Oxoid) containing 6% (v/v) horse blood. *E.coli* strains were grown in Luria broth (28) and plated on Luria broth containing 1.5% (w/v) agar. If required, antibiotics were added to the plates at the following concentrations: spectinomycin: 100 ug/ml for *S. suis* and 50 ug/ml for *E. coli* and ampicillin, 50 ug/ml.

Serotyping. The *S.suis* strains were serotypes by the slide agglutination test with serotype-specific antibodies (44).

DNA techniques. Routine DNA manipulations were performed as described by Sambrook et al. (36).

Alkaline phosphatase activity. To screen for PhoA fusions in *E.coli*, plasmid libraries were constructed. Therefore, chromosomal DNA of *S. suis* type 2 was digested with *Alu*I. The 300-500-bp fragments were ligated to *Sma*I-digested pPHOS2. Ligation mixtures were transformed to the PhoA$^-$ *E. coli* strain CC118. Transformants were plated on LB media supplemented with 5-Bromo-4-chloro-3-indolylfosfaat (BCIP, 50 ug/ml, Boehringer, Mannheim, Germany). Blue colonies were purified on fresh LB/BCIP plates to verify the blue phenotype.

DNA sequence analysis. DNA sequences were determined on a 373A DNA Sequencing System (Applied Biosystems, Warrington, GB). Samples were prepared by use of a ABI/PRISM dye terminator cycle sequencing ready reaction kit (Applied Biosystems). Sequencing data were assembled and analyzed using the MacMollyTetra program. Custom-made sequencing primers were purchased from Life Technologies. Hydrophobic stretches within

proteins were predicted by the method of Klein et al. (17). The BLAST program available on Netscape Navigator™ was used to search for protein sequences related to the deduced amino acid sequences.

Construction of gene-specific knock-out mutants of *S. suis*. To construct the mutant strains 10cpsB and 10cpsEF we *electrotransformed the pathogenic serotype 2 strain* 10 (45, 49) of *S. suis* with pCPS11 and pCPS28 respectively. In these plasmids the *cps*B and *cps*EF genes were disturbed by the insertion of a spectinomycin-resistance gene. To create pCPS11 the internal 400 bp *Pst*I-*Bam*HI fragment of the *cps*B gene in pCPS7 was replaced by the Spc^R gene. For this purpose pCPS7 was digested with *Pst*I and *Bam*HI and ligated to the 1,200-bp *Pst*I-*Bam*HI fragment, containing the Spc^R gen, from pIC-spc. To construct pCPS28 we have used pIC20R. In this plasmid we inserted the *KpnI-SalI fragment from* pCPS17 (resulting in pCPS25) and the *Xba*I-*Cla*I fragment from pCPS20 (resulting in pCPS27). pCPS27 was digested with *Pst*I and *Xho*I and ligated to the 1,200-bp *Pst*I-*Xho*I fragment, containing the Spc^R gene of pIC-spc. The electrotransformation to *S. suis* was carried out as described before (38).

Southern blotting and hybridization. Chromosomal DNA was isolated as described by Sambrook *et al.* (36). DNA fragments were separated on 0.8% agarose gels and transferred to Zeta-Probe GT membranes (Bio-Rad) as described by Sambrook *et al.* (36). DNA probes were labelled with [($-^{32}$P]dCTP (3000 Ci mmol^{-1}; Amersham) by use of a random primed labelling kit (Boehringer). The DNA on the blots was hybridized at 65°C with appropriate DNA probes as recommended by the supplier of the Zeta-Probe membranes. After hybridization, the membranes were washed twice with a solution of 40 mM sodium phosphate, pH 7.2, 1 mM EDTA , 5% SDS for 30 min at 65°C and twice with a solution of 40 mM sodium phosphate, pH 7.2, 1 mM EDTA, 1% SDS for 30 min at 65°C.

PCR. The primers used in the *cps2J* PCR correspond to the positions 13791-13813 and 14465-14443 in the *S. suis cps2*

locus. The sequences were: 5'-CAAACGCAAGGAATTACGGTATC-3' and 5'-GAGTATCTAAAGAATGCCTATTG-3'. The primers used for the *cps1I* PCR correspond to the positions 4398-4417 and 4839-4821 in the *S. suis cps1* sequence. The sequences were: 5'-GGCGGTCTAGCAGATGCTCG-3' and 5'-GCGAACTGTTAGCAATGAC-3'. The primers used in the *cps9H* PCR correspond to the positions 4406-4126 and 4494-4475 in the *S. suis cps9* sequence. The sequences were: 5'-GGCTACATATAATGGAAGCCC3' and 5'-CGGAAGTATCTGGGCTACTG-3'.

Construction of gene-specific knock-out mutants of *S. suis*. To construct the mutant strains 10cpsB and 10cpsEF we electrotransformed the pathogenic serotype 2 strain 10 of *S. suis* with pCPS11 and pCPS28 respectively. In these plasmids the *cps*B and *cps*EF genes were disturbed by the insertion of a spectinomycin-resistance gene. To create pCPS11 the internal 400 bp *Pst*I-*Bam*HI fragment of the *cps*B gene in pCPS7 was replaced by the Spc^R gene. For this purpose pCPS7 was digested with *Pst*I and *Bam*HI and ligated to the 1,200-bp *Pst*I-*Bam*HI fragment, containing the Spc^R gen, from pIC-spc. To construct pCPS28 we have used pIC20R. In this plasmid we inserted the *Kpn*I-*Sal*I fragment from pCPS17 (resulting in pCPS25) and the *Xba*I-*Cla*I fragment from pCPS20 (resulting in pCPS27). pCPS27 was digested with *Pst*I and *Xho*I and ligated to the 1,200-bp *Pst*I-*Xho*I fragment, containing the Spc^R gene of pIC-spc. The electrotransformation to *S. suis* was carried out as described before (38).

Phagocytosis assay. Phagocytosis assays were performed as described by Leij *et al.* (23). Briefly, to opsonize the cells, 10^7 *S. suis* cells were incubated with 6% SPF-pig serum for 30 min at 37°C in a head-over-head rotor at 6 rpm. 10^7 AM and 10^7 opsonized *S. suis* cells were combined and incubated at 37°C under continuous rotation at 6 rpm. At 0, 30, 60 and 90 min, 1-ml samples were collected and mixed with 4 ml of ice-cold EMEM to stop phagocytosis. Phagocytes were removed by centrifugation for 4 min at 110 x g and 4°C. The number of colony forming units (CFU) in the supernatants was determined. Control

experiments were carried out simultaneously by combining 10^7 opsonized *S. suis* cells with EMEM (without AM).

Killing assays. AM (10^7/ml) and opsonized *S. suis* cells (10^7/ml) were mixed 1 : 1 and incubated for 10 min at 37°C under continuous rotation at 6 rpm. Ice-cold EMEM was added to stop further phagocytosis and killing. To remove extracellular *S. suis* cells, phagocytes were washed twice (4 min, 110 x g, 4°C) and resuspended in 5 ml EMEM containing 6% SPF serum. The tubes were incubated at 37°C under rotation at 6 rpm. After 0, 15, 30, 60 and 90 min, samples were collected and mixed with ice-cold EMEM to stop further killing. The samples were centrifuged for 4 min at 110 x g at 4°C and the phagocytic cells were lysed in EMEM containing 1% saponine for 20 min at room temperature. The number of CFU in the suspensions was determined.

Pigs. Germfree pigs, cross-breeds of Great Yorkshire and Dutch landrace, were obtained from sows by caesarian sections. The surgery was performed in sterile flexible film isolators. Pigs were allotted to groups, each consisting of 4 pigs, and were housed in sterile stainless steel incubators.

Experimental infections. Pigs were inoculated intranasally with *S. suis* type 2 as described before. To predispose the pigs for infection with *S. suis*, five-day old pigs were inoculated intranasally with about 10^7 CFU of *Bordetella bronchiseptica* strain 92932. Two days later the pigs were inoculated intranasally with *S. suis* type 2 (10^6 CFU). Pigs were monitored twice daily for clinical signs of disease, such as fever, nervous signs and lameness. Blood samples were collected three times a week from each pig. White blood cells were counted with a cell counter. To monitor infection with *S. suis* and *B. bronchiseptica* and to check for absence of contaminants, we collected swabs of nasopharynx and feces daily. The swabs were plated directly onto Columbia agar containing 6% horse blood. After three weeks the pigs were killed and examined for pathological changes. Tissue specimens from the central nervous system, serosae, and joints were examined bacteriologically and

histologically as described before (45, 49). Colonization of the serosae was scored positively when *S. suis* was isolated from the pericardium, thoracal pleura or the peritoneum. Colonization of the joints was scored positively when *S. suis* was isolated from one or more joints (12 joints per animal were scored).

Vaccination and challenge

One week old pigs were vaccinated intravenously with a dosage of 106 cfu of the S. suis strains 10cpsEF or 10cpsB. Three weeks later the pigs were challenged intravenously with the pathogenic serotype 2 strain 10 (107 cfu). Disease monitoring, haematological, serological and bacteriological examinations as well as post-mortum examinations were as described before under experimental infections.

Electron Microscopy. Bacteria were prepared for electron microscopy as described by Wagenaar et al. (50). Shortly, bacteria were mixed with agarose MP (Boehringer) of 37° C to a concentration of 0.7%. The mixture was immediately cooled on ice. Upon gelifying, samples were cut into 1 to 1.5 mm slices and incubated in a fixative containing 0.8% glutaraldehyde and 0.8% osmiumtetraoxide. Subsequently, the samples were fixed and stained with uranyl acetate by microwave stimulation, dehydrated and imbedded in eponaraldite resin. Ultra-thin sections were counterstained with lead citrate and examined with a Philips CM 10 electron microscope at 80 kV.

Isolation of porcine alveolar macrophages (AM). Porcine AM were obtained from the lungs of specific pathogen free (SPF) pigs. Lung lavage samples were collected as described by van Leengoed *et al.* (43). Cells were suspended in EMEM containing 6% (v/v) SPF-pig serum and adjusted to 10^7 cells per ml.

RESULTS

Identification of the *cps* locus.

The *cps* locus of *S.suis* type 2 was identified by making use of a strategy developed for the genetic identification of exported proteins (13, 31). In this system we made use of a plasmid (pPHOS2) containing a truncated alkaline phosphatase gene (13). The gene lacked the promoter sequence, the translational start site and the signal sequence. The truncated gene is preceded by a unique *Sma*I restriction site. Chromosomal DNA of *S. suis* type 2, digested with *Alu*I, was randomly cloned in this restriction site. Because translocation of PhoA across the cytoplasmic membrane of *E. coli* is required for enzymatic activity, the system can be used to select for *S. suis* fragments containing a promoter sequence, a translational start site and a functional signal sequence. Among 560 individual *E. coli* clones tested, 16 displayed a dark blue phenotype when plated on media containing BCIP. DNA sequence analysis of the inserts from several of these plasmids were performed (results not shown) and the deduced amino acid sequences were analyzed. The hydrophobicity profile of one of the clones (pPHOS7, results not shown) showed that the N-terminal part of the sequence resembled the characteristics of a typical signal peptide: a short hydrophilic N-terminal region is followed by a hydrophobic region of 38 amino acids. These data indicate that the phoA system was successfully used for the selection of *S. suis* genes encoding exported proteins. Moreover, the sequences were analyzed for similarities present in the databases. The sequence of pPHOS7 showed a high similarity (37% identity) with the protein encoded by the *cps14C* gene of *Streptococcus pneumoniae* (19). This strongly suggests that pPHOS7 contains a part of the *cps* operon of *S. suis* type 2.

Cloning of the flanking *cps* genes. In order to clone the flanking *cps* genes of *S. suis* type 2 the insert of pPHOS7 was used as a probe to identify chromosomal DNA fragments which contain flanking *cps* genes. A 6-kb *Hind*III fragment was

identified and cloned in pKUN19. This yielded clone pCPS6 (Fig. 1C). Sequence analysis of the insert of pCPS6 revealed that pCPS6 most probably contained the 5'-end of the *cps* locus, but still lacked the 3'-end. Therefore, sequences of the 3' -end of pCPS6 were in turn used as a probe to identify chromosomal fragments containing *cps* sequences located further downstream. These fragments were also cloned in pKUN19, resulting in pCPS17. Using the same system of chromosomal walking we subsequently generated the plasmid pCPS18, pCPS20, pCPS23 and pCPS26, containing downstream *cps* sequences.

Analysis of the cps operon. The complete nucleotide sequence of the cloned fragments was determined (figure 4). Examination of the compiled sequence revealed the presence of at least 13 potential open reading frame (Orfs), which were designated as Orf 2Y, Orf2X and Cps2A-Cps2K (Fig. 1A). Moreover, a 14th, incomplete, Orf (Orf 2Z) was located at the 5'-end of the sequence. Two potential promoter sequences were identified. One was located 313 bp (locations 1885-1865 and 1884-1889) upstream of Orf2X. The other potential promoter sequence was located 68 bp upstream of Orf2Y (locations 2241-2236 and 2216-2211). Orf2Y is expressed in opposite orientation. Between Orfs 2Y and 2Z the sequence contained a potential stem-loop structure, which could act as a transcription terminator. Each Orf is preceded by a ribosome-binding site and the majority of the Orfs are very closely linked. The only significant intergenic gap was found between Cps2G and Cps2H (389 nucleotides). However, no obvious promoter sequences or potential stem-loop structures were found in this region. These data suggest that Orf2X and Cps2A-Cps2K are arranged as an operon.

An overview of all Orfs with their properties is shown in Table 2. The majority of the predicted gene products is related to proteins involved in polysaccharide biosynthesis. Orf2Z showed some similarity with the YitS protein of *Bacillus subtilis*. YitS was identified during the sequence analysis of the complete genome of *B. subtilis*. The function of the protein

is unknown.

Orf2Y showed similarity with YcxD protein of *B. subtilis* (53). Based on the similarity between YcxD and MocR of *Rhizobium meliloti* (33), YcxD was suggested to be a regulatory protein.

Orf2X showed similarity with the hypothetical YAAA proteins of *Haemophilus influenzae* and *E. coli*. The function of these proteins is unknown.

The gene products encoded by the *cps2A*, *cps2B*, *cps2C* and *cps2D* genes showed approximate similarity with the CpsA, CpsC, CpsD and CpsB proteins of several serotypes of *Streptococcus pneumoniae* (19), respectively. This suggest similar functions for these proteins. Hence, Cps2A may have a role in the regulation of the capsular polysaccharide synthesis. Cps2B and Cps2C could be involved in the chain length determination of the type 2 capsule and Cps2C can play an additional role in the export of the polysaccharide. The Cps2D protein of *S. suis* is related to the CpsB protein of *S. pneumoniae* and to proteins encoded by genes of several other Gram-positive bacteria involved in polysaccharide or exopolysaccharide synthesis, but their function is unknown (19).

The protein encoded by *cps2E* gene showed similarity to several bacterial proteins with glycosyl transferase activities: Cps14E and Cps19fE of *S. pneumoniae* serotypes 14 and 19F (18, 19, 29), CpsE of *Streptococcus salvarius* (X94980) and CpsD of *Streptococcus agalactiae* (34). Recently, Kolkman *et al.* (18) showed that Cps14E is a glucosyl-1-phosphate transferase that links glucose to a lipid carrier, the first step in the biosynthesis of the S. *pneumoniae* type 14 repeating unit. Based on these data a similar function may be fulfilled by Cps2E of *S. suis* .

The protein encoded by the *cps2F* gene showed similarity to the protein encoded by the *rfbU* gene of *Salmonella enteritica*. (25). This similarity is most pronounced in the C-terminal regions of these proteins. The *rfbU* gene was shown to encoded mannosyltransferase activity (25).

The *cps2G* gene encoded a protein that showed moderate similarity with the *rfbF* gene product of *Campylobacter hyoilei* (22), the *epsF* gene product of *S. thermophilus* (40) and the *capM* gene product of *S. aureus* (24). On the basis of similarity the *rfbF, epsF* and *capM* genes are suggested to encoded galactosyltransferase activities. Hence, a similar glycosyl transferase activity could be fulfilled by the *cps2G* gene product.

The *cps2H* gene encodes a protein that is similar to the N-terminal region of the *lgtD* gene product of *Haemophilus influenzae* (U32768). Moreover, the hydrophobicity plots of Cps2H and LgtD looked very similar in these regions (data not shown). Based on sequence similarity the *lgtD* gene product was suggested to have glycosyl transferase activity (U32768).

The gene product encoded by the *cps2I* gene showed some similarity with a protein of *Actinobacillus actinomycetemcomitans* (AB002668). This protein is part of the gene cluster responsible for the serotype-b-specific antigen of *A. actimycetemcomitans*. The function of the protein is unknown.

The gene products encoded by the *cps2J* and *cps2K* genes showed significant similarities to the Cps14J protein of *S. pneumoniae*. The *cps14J* gene of *S. pneumoniae* was shown to encode a ß-1,4-galactosyltransferase activity. In *S. pneumoniae* CpsJ is responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide (20). Even some similarity was found between Cps2J and Cps2K (Fig. 2, 25.5% similarity). This similarity was most pronounced in the N-terminal regions of the proteins. Recently, two small conserved regions were identified in the N-terminus of Cps14J and Cps14I and their homologues (20). These regions were predicted to be important for catalytic activity. Both regions, DXS and DXDD (Fig. 2), were also found in Cps2J and Cps2K.

Distribution of the *cps2* genes in other *S. suis* serotypes. To examine the relationship between the *cps2* genes and *cps* genes in the other *S. suis* serotypes, we performed cross-hybridization experiments. DNA fragments of the individual *cps2* genes were amplified by PCR, labelled with ^{32}P, and used to probe Southern blots of chromosomal DNA of the reference strains of the 35 different *S. suis* serotypes. Large variation in the hybridization patterns were observed (Table 4). As a positive control we used a probe specific for 16S rRNA. The 16S rRNA probe hybridized with all serotypes tested. However, none of the other genes tested were common in all serotypes. Based on the genetic organization of the genes we previously suggested that *orfX* and *cpsA-cpsK* genes are part of one operon and that the protein encoded by these genes are all involved in polysaccharide biosynthesis. OrfY and OrfZ are not a part of this operon, and their role in the polysaccharide biosynthesis is unclear. Based on sequence similarity data, OrfY may be involved in regulation of the *cps2* genes. OrfZ is proposed to be unrelated to polysaccharide biosynthesis. Probes specific for the *orfZ, orfY, orfX, cpsA, cpsB, cpsC* and *cpsD* genes hybridized with most other serotypes. This suggests that the protein encoded by these genes are not type-specific, but may perform more common functions in biosynthesis of the capsular polysaccharide. This confirms previous data which showed that the *cps2A-cps2D* genes showed strong similarity to *cps* genes of several serotype of *Streptococcus pneumoniae*. Based on this similarity Cps2A is possibly a regulatory protein, whereas Cps2B and Cps2C may play a role in length determination and export of polysaccharide. The *cps2E* gene hybridized with DNA of serotypes 1, 2, 14 and 1/2. The *cps2E* gene showed a strong similarity to the *cps14E* gene of *S. pneumoniae* (18). This enzyme was shown to have a glucosyl-1-phosphate activity and catalyzed the transfer of glucose to a lipid carrier (18). These data indicate that a glycosyltransferase closely related to Cps14E may be responsible for the first step in the biosynthesis of

polysaccharide in the *S. suis* serotypes 1, 2, 14 and 1/2. The *cps2F*, *cps2G*, *cps2H*, *cps2I* and *cps2J* genes hybridized with chromosomal DNA of serotypes 2 and 1/2 only. The *cps2G* gene showed an additional weak hybridization signal with DNA of serotype 34. In agglutination tests serotype 1/2 showed agglutination with sera specific for serotype 2 as well as with sera specific for serotype 1. This suggests that serotype 1/2 shares antigenic determinants with both types 1 and 2. The hybridization data confirmed these data. All putative glycosyltransferases present in serotype 2 are also present in serotype 1/2. The *cps2K* gene showed a similar hybridization pattern as the *cps2E* gene. Hybridization was observed with DNA of serotypes 1, 2, 14 and 1/2. Taken together these hybridization data show that the *cps2* gene cluster can be divided in three regions: a central region containing the type-specific genes is flanked by two regions containing common genes for various serotypes.

Cloning of the type-specific *cps* genes of serotypes 1 and 9.

To clone the type-specific *cps* genes of *S. suis* serotype 1 we used the *cps2E* gene as a probe to identify chromosomal DNA fragments of type 1 which contain flanking *cps* genes. A 5 kb *Eco*RV fragment was identified and cloned in pKUN19. This yielded pCPS1-1 (Fig. 1B). This fragment was in turn used as a probe to identify an overlapping 2.2 kb *Hin*dIII fragment. pKUN19 containing this *Hin*dIII fragment was designated pCPS1-2. The same strategy was followed to identify and clone the type-specific *cps* genes of serotype 9. In this case, we used the *cps2D* gene as a probe. A 0.8 kb *Hin*dIII-*Xba*I fragment was identified and cloned, yielding pCPS9-1 (Fig. 1C). This fragment was in turn used as a probe to identify a 4 kb *Xba*I fragment. pKUN19 containing this 4 kb *Xba*I fragment was designated pCPS9-2.

Analysis of the cloned *cps1* genes. The complete nucleotide sequence of the inserts of pCPS1-1 and pCPS1-2 was determined (figure 5). Examination of the sequence revealed the presence of five complete and two incomplete Orfs (Fig.1B). Each Orf is preceded by a ribosome-binding site. In accord with data obtained for the *cps2* genes of serotype 2, the majority of the Orfs is very closely linked. The only significant gap (718 bp) was found between Cps1G and Cps1H. No obvious promoter sequences or potential stem-loop structures could be found in this region. This suggests that, as in serotype 2, the *cps* genes in serotype 1 are arranged in an operon.

An overview of the Orfs and their properties in shown in Table 2. As expected on the basis of the hybridization data (Table 4), the protein encoded by the *cps1E* gene was related to Cps2E of *S. suis* type 2 (identity of 86%). The fragment cloned in pCPS1-1 lacked the coding region for the first 7 amino acids of the *cps1E* gene.

The protein encoded by the *cps1F* and *cps1G* genes showed strong similarity to the Cps14F and Cps14G proteins of *Streptococcus pneumoniae* serotype 14, respectively (20). The function of the Cps14F is not completely clear, but it has been suggested that Cps14F can enhance role in glycosyltransferase activity. The *cps14G* gene of *S. pneumoniae* was shown to encode ß-1,4-galactosyltransferase activity. In *S. pneumoniae* type 14 this activity is required for the second step in the biosynthesis of the oligosaccharide subunit (20). Based on the similarity data found similar glycosyltransferase and enhancing activities are suggested for the *cps 1G* and *cps1F* genes of *S. suis* type 1.

The protein encoded by the *cps1H* gene showed similarity to the Cps14H protein of *S. pneumoniae* (20). Based on sequence similarity Cps14H was proposed to be the polysaccharide polymerase (20).

The protein encoded by the *cps1I* gene showed some similarity with the Cps14J protein of *S. pneumoniae* (19). The *cps14J* gene was shown to encode a ß-1,4-galactosyltransferase

activity, responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide.

Between Cps1G and Cps1H a gap of 718 bp was found. This region revealed three small Orfs. The three Orfs were expressed in three different reading frames and were not preceded by potential ribosome binding sites, nor contained potential start sites. However, the three potential gene products encoded by this region showed some similarity with three successive regions of the C-terminal part of the EpsK protein of *Streptococcus thermophilus* (27% identity, 40). The region related to the first 82 amino acids is lacking.

Analysis of the cloned *cps9* genes. We also determined the complete nucleotide sequence of the inserts of pCPS9-1 and pCPS9-2 (figure 6). Examination of the sequence revealed the presence of three complete and two incomplete Orfs (Fig.1C). As in serotypes 1 and 2, all Orfs are preceded by a ribosome-binding site and are very closely coupled. As suggested by the hybridization data (Table 4) the Cps2D and Cps9D proteins were highly related (Table 2). Based on sequence comparisons pCPS9-1 lacked the first 27 amino acids of the Cps9D protein.

The protein encoded by the *cps9E* gene showed some similarity with the CapD protein of *Staphylococcus aureus* serotype 1 (24). Based on sequence similarity data the Cap1D protein was suggested to be an epimerase or a dehydratase involved in the synthesis of N-acetylfructosamine or N-acetylgalactosamine (63).

Cps9F showed some similarity to the CapM proteins of *S. aureus* serotypes 5 and 8 (61, 64, 65). Based on sequence similarity data Cap5M and Cap8M are proposed to be glycosyltransferases (63).

The protein encoded by the *cps9G* gene showed some similarity with a protein of *Actinobacillus actinomycetemcomitans* (AB002668_4). This protein is part of a gene cluster responsible for the serotype-b specific antigens

of *Actinobacillus actinomycetemcomitans* . The function of the protein is unknown.

The protein encoded by the *cps9H* gene showed some similarity with the *rfbB* gene of *Yersinia enterolitica* (68). The RfbB protein was shown to be essential for O-antigen synthesis, but the function of the protein in the synthesis of the O:3 lipopolysaccharide is unknown.

Serotype 1 and serotype 9 specific *cps* genes. To determine whether the cloned fragments in pCPS1-1, pCPS1-2, pCPS9-1 and pCPS9-2 contained the type-specific genes for serotype 1 and 9, respectively, cross hybridization experiments were performed. DNA fragments of the individual *cps1* and *cps9* genes were amplified by PCR, labelled with ^{32}P, and used to probe Southern blots of chromosomal DNA of the reference strains of the 35 different *S. suis* serotypes. The results are shown in Table 5. Based on the data obtained with the *cps2E* probe (Table 4), the *cps1E* probe was expected to hybridize with chromosomal DNA of *S. suis* serotypes 1,2, 14, 27 and 1/2. The *cps1H, cps9E* and *cps9F* probes hybridized with most other serotypes. However, the *cps1F* and *cps1G* and cps1I probes hybridized with chromosomal DNA of serotypes 1 and 14 only. The *cps9G* and *cps9H* probe hybridized with serotype 9 only. These data suggest that the *cps9G* and cps9H probes are specific for serotype 9 and therefore could be useful tools for the development of rapid and sensitive diagnostic tests for *S. suis* type 9 infections.

Type specific PCR. So far, the probes were tested on the 35 different reference strains only. To test the diagnostic value of the type-specific *cps* probes further, several other *S. suis* serotype 1, 2, 1/2, 9 and 14 strains were used. Moreover, since a PCR based method would be even more rapid and sensitive than a hybridization test, we tested whether we could use a PCR for the serotyping of the *S. suis* strains. The

oligonucleotide primer sets were chosen within the *cps2J*, *cps1I* and *cps9H* genes. Amplified fragments of 675 bp, 380 bp and 390 bp were expected respectively. The results show that 675 bp fragments were amplified on type 2 and 1/2 strains using *cps2J* primers; 380 bp fragments were amplified on type 1 and 14 strains using *cps1I* primers and 390 bp fragments were amplified on type 9 strains using *cps9H* primers.

Construction of mutants impaired in capsule production. To evaluate the role of the capsule of *S. suis* type 2 in the pathogenesis, we constructed two isogenic mutants in which capsule production was disturbed. To construct mutant 10cpsB, pCPS11 was used. In this plasmid a part of the *cps2B* gene was replaced by the spectinomycin-resistance gene. To construct mutant strain 10cpsEF the plasmid pCPS28 was used. In pCPS28 the 3'-end of *cps2E* gene as well as the 5'-end of *cps2F* gene were replaced by the spectinomycin-resistance gene. pCPS11 and pCPS28 were used to electrotransform strain 10 of *S. suis* type 2 and spectinomycin-resistant colonies were selected. Southern blotting and hybridization experiments were used to select double cross over integration events (results not shown). To test whether the capsular structure of the strains 10cpsB and 10cpsEF was disturbed, we used a slide agglutination test using a suspension of the mutant strains in hyperimmune anti-*S. suis* type 2 serum (44). The results showed that even in the absence of serotype specific antisera, the bacteria agglutinated. This indicates that in the mutant strains the capsular structure was disturbed. To confirm this, thin sections of wild type and mutant strains were compared by electron microscopy. The results showed that compared to the wild type (Fig. 3A) the amount of capsule produced by the mutant strains was greatly reduced (Figs. 3B and 3C). Almost no capsular material could be detected on the surface of the mutant strains.

Capsular mutants are sensitive to phagocytosis and killing by porcine alveolar macrophages (PAM).

The capsular mutants were tested for their ability to resist phagocytosis by PAM in the presence of porcine SPF serum. The wild type strain 10 seemed to be resistant to phagocytosis under these conditions (Fig. 4A). In contrast, the mutant strains were efficiently ingested by macrophages (Fig. 4A). After 90 min. more than 99.7% (strain 10cpsB) and 99.8% (strain 10cpsEF) of the mutant cells were ingested by the macrophages. Moreover, as shown in Fig. 4B the ingested strains were efficiently killed by the macrophages. 90-98% of all ingested cells were killed within 90 min. No differences could be observed between wild type and mutant strains. These data indicate that the capsule of *S. suis* type 2 efficiently protects the organism from uptake by macrophages *in vitro*.

Capsular mutants are less virulent for germfree piglets. The virulence properties of the wild-type and mutant strains were tested after experimental infection of newborn germfree pigs (45, 49). Table 1 shows that specific and nonspecific signs of disease could be observed in all pigs inoculated with the wild type strain. Moreover, all pigs inoculated with the wild type strain died during the course of the experiment or were killed because of serious illness or nervous disorders (Table 3). In contrast, the pigs inoculated with strains 10cpsB and 10cpsEF showed no specific signs of disease and all pigs survived until the end of the experiment. The temperature of the pigs inoculated with the wild type strain increased 2 days after inoculation and remained high until day 5 (Table 3). The temperature of the pigs inoculated with the mutant strains sometimes exceeded the 40^{o}C, however, we could observe significant differences in the fever index [i.e % of observations in an experimental group during which pigs showed fever (>40^{o}C)] between pigs inoculated with wild type and mutant strains. All pigs showed increased numbers of polymorphonuclear leucocytes (PMLs) (>10 x 10^{9} PMLs per litre)

(Table 3). However, in pigs inoculated with the mutant strains the percentage of samples with increased numbers of PMLs was considerably lower. *S. suis* strains and *B. bronchiseptica* could be isolated from the nasopharynx and feces swab samples of all pigs from 1 day post-infection until the end of the experiment (Table 3). Postmortem, the wild type strain could frequently be isolated from the central nervous system (CNS), kidney, heart, liver , spleen, serosae, joints and tonsils. Mutant strains could easily be recovered form the tonsils, but were never recovered from the kidney, liver or spleen. Interestingly, low numbers of the mutant strains were isolated from the CNS, the serosae, the joints,the lungs and the heart. Taken together, these data strongly indicated that mutant *S. suis* strains, impaired in capsule production, are not virulent for young germfree pigs.

We describe the identification and the molecular characterisation of the *cps* locus, involved in the capsular polysaccharide biosynthesis, of *S. suis* Most of the genes seemed to belong to a single transcriptional unit, suggesting a co-ordinate control of these genes. We assign functions to most of the gene products. We thereby identified regions involved in regulation (Cps2A), chain length determination (Cps2B, C), export (Cps2C) and biosynthesis (Cps2E, F, G, H, J, K). The region involved in biosynthesis is located at the centre of the gene cluster and is flanked by two regions containing genes with more common functions. The incomplete *orf2Z* gene was located at the 5'-end of the cloned fragment. Orf2Z showed some similarity with the YitS protein of *B.subtilis*. However, because the function of the YitS protein is unknown this did not give us any information about the possible function of Orf2Z. Because the *orf2Z* gene is not a part of the *cps* operon, a role of this gene in polysaccharide biosynthesis is not expected. The Orf2Y protein showed some similarity with the YcxD protein of *B.subtilis* (53).The YcxD protein was suggested to be a regulatory protein. Similarly, Orf2Y may be involved in the regulation of polysaccharide biosynthesis. The Orf2X

protein showed similarity with the YAAA proteins of *H. influenzae* and *E. coli*. The function of these proteins is unknown. In *S. suis* type 2 the *orf2X* gene seemed to be the first gene in the *cps2* operon. This suggests a role of Orf2X in the polysaccharide biosynthesis. In *H. influenzae* and *E. coli*, however, these proteins are not associated with capsular gene clusters. The analysis of isogenic mutants impaired in the expression of Orf2X should give more insight in the presumed role of Orf2X in the polysaccharide biosynthesis of *S. suis* type 2.

The gene products encoded by the *cps2E, cps2F, cps2G, cps2H, cps2J* and *cps2K* genes showed little similarity with glycosyltransferases of several Gram-positive or Gram-negative bacteria (18, 19, 20, 22, 25). The *cps2E* gene product shows some similarity with the Cps14E protein of *S. pneumoniae* (18, 19). Cps14E is a glucosyl-1-phosphate transferase that links glucose to a lipid carrier (18). In *S. pneumoniae* this is the first step in the biosynthesis of the oligosaccharide repeating unit. The structure of the *S. suis* serotype 2 capsule contains glucose, galactose, rhamnose, N-acetyl glucoseamine and sialic acid in a ratio of 3:1:1:1:1 (7). Based on these data we conclude that Cps2E of *S. suis* has glucosyltransferase activity, and is involved in the linkage of the first sugar to the lipid carrier.

The C-terminal region of the *cps2F* gene product showed some similarity with the RfbU of *Salmonella enteritica*. RfbU was shown to have mannosyltransferase activity (24). Because mannosyl is not a component of the *S. suis* type 2 polysaccharide a mannosyltransferase activity is not expected in this organism. Nevertheless, *cps2F* encodes a glycosyltransferase with another sugar specificity.

Cps2G showed moderate similarity to a family of gene products suggested to encode galactosyltransferase activities (22, 24, 40). Hence a similar activity is shown for Cps2G.

Cps2H showed some similarity with LgtD of *H. influenzae* (U32768). Because LgtD was proposed to have glycosyltransferase

activity , a similar activity is fulfilled by Cps2H.

Cps2J and Cps2K showed similarity to Cps14J of *S. pneumoniae* (20). *Cps2J showed similarity with Cps14I of S. pneumoniae as* well. Cps14I was shown to have N-acetyl glucosaminyltransferase activity, whereas Cps14J has a ß-1,4-galactosyltransferase activity (20). In *S. pneumoniae* Cps14I is responsible for the addition of the third sugar and *Cps14J* for the addition of the last sugar in the synthesis of the type 14 repeating unit (20). Because the capsule of *S. suis* type 2 contains galactose as well as N-acetyl glucosamine components, galactosyltransferase as well as N-acetyl glucoaminyltransferase activities could be envisaged for the *cps2J* and *cps2K* gene products, respectively. As was observed for Cps14I and Cps14J, the N-termini of Cps2J and Cps2K showed a significant degree of sequence similarity. Within the N-terminal domains of Cps14I and Cps14J, two small regions were identified, which were also conserved in several other glycosyltransferases (22). Within these two regions, two Asp residues were proposed to be important for catalytic activity. The two conserved regions, DXS and DXDD, were also found in Cps2J and Cps2K.

The function of Cps2I remains unclear. Cps2I showed some similarity with a protein of *A. actinomycetemcomitans*. Although this protein part is of the gene cluster responsible for the serotype-B-specific antigens, the function of the protein is unknown.

We further describe the identification and characterization of the *cps* genes specific for *S. suis* serotypes 1, 2 and 9. After the entire *cps2* locus of *S. suis* serotype 2 was cloned and characterized, functions for most of the *cps2* gene products could be assigned by sequence homologies. Based on these data the glycosyltransferase activities, required for type specificity, could be located in the centre of the operon. Cross-hybridization experiments, using the individual *cps2* genes as probes on chromosomal DNAs of the 35 different serotypes, confirmed this idea. The regions containing the

type-specific genes of serotypes 1 and 9 could be cloned and characterized, showing that an identical genetic organization of the *cps* operons of other *S. suis* serotypes exists. The *cps1E, cps1F, cps1G, cps1H,* and *cps1I* genes revealed a striking similarity with *cps14 E, cps14F, cps14G, cps14H* and *cps14J* genes of *S. pneumoniae*. Interestingly, *S. pneumoniae* serotype 14 is the serotype most commonly associated with pneumococcal infections in young children (54), whereas *S. suis* serotype 1 strains are most commonly isolated from piglets younger than 8 weeks (46). In *S. pneumoniae* the *cps14E, cps14G, cps14I* and *cps14J* encode the glycosyltransferases required for the synthesis of the type 14 tetrameric repeating unit, showing that the *cps1E, cps1G* and *cps1I* genes encoded glycosyltransferases. The precise functions of these genes as well as the substrate specificities of the enzymes can be established. In *S. pneumoniae* the *cps14E* gene was shown to encode a glucosyl-1-phosphate transferase catalyzing the transfer of glucose to a lipid carrier. Moreover, *cpsE*-like genes were found in *S. pneumoniae* serotypes 9N, 13, 14, 15B, 15C, 18F, 18A and 19F (60). *CpsE* mutants were constructed in the serotypes 9N, 13 , 14 and 15B. All mutant strains lacked glucosyltransferase activity (60). Moreover, in all these *S. pneumoniae* serotypes the *cpsE* gene seemed to be responsible for the addition of glucose to the lipid carrier. Based on these data we suggest that in *S. suis* type 1 the *cps1E* gene may fulfil a similar function. The structure of the *S. suis* type 1 capsule is unknown, but it is composed of glucose, galactose, N-acetyl glucosamine, N-acetyl galactosamine and sialic acid in a ratio of 1: 2.4: 1: 1:1.4 (5). Therefore a role of a *cpsE*-like glucosyltransferase activity can easily be envisaged. *CpsE* like sequences were also found in serotypes 2, 1/2 and 14.

For polysaccharide biosynthesis in *S. pneumoniae* type 14, transfer of the second sugar of the repeating unit to the first lipid-linked sugar is performed by the gene products of *cps14F* and *cps14G* (20). Similar to Cps14F and Cps14G, the *S.*

suis type 1 proteins Cps1F and Cps1G may act as one glycosyltransferase performing the same reaction. Cps14F and Cps14G of *S. pneumoniae* showed similarity to the N-terminal half and C-terminal half of the SpsK protein of *Sphingomonas* (20, 67), respectively. This suggests a combined function for both proteins. Moreover, *cps14F* and *cps14G* like sequences were found in several serotypes of *S. pneumoniae* and these genes always seemed to exist together (60). The same was observed for *S. suis type* 1. The *cps1F* and *cps1G* probes hybridized with type 1 and type 14 strains.

According to the similarity found between the *cps1H* gene and the *cps14H* gene of *S. pneumoniae* (20), *cps1H* is expected to encode a polysaccharide polymerase.

The protein encoded by the *cps1I* gene showed some similarity with the Cps14J protein of *S. pneumoniae* (19). The *cps14J* gene was shown to encode a ß-1,4-galactosyltransferase activity, responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide. In *S. suis* type 2 the proteins encoded by the *cps2J* and *cps2K* genes showed similarity to the Cps14J protein. However, no significant homologies were found between Cps2J, Cps2K and Cps1I. In the N-terminal regions of Cps14J and Cps14I two small conserved regions, DXS and DXDD, were identified (19). These regions seemed to be important for catalytic activity (13). At the same positions in the sequence Cps2I contained the regions DXS and DXED.

In the region between Cps1G and Cps1H three small Orfs were identified. Since the Orfs were expressed in three different reading frames, and did not contain potential start sites, expression is not expected. However, the three potential gene products encoded by this region showed some similarity with three successive regions of the C-terminal part of the EpsK protein of *Streptococcus thermophilus* (27% identity, 40). The region related to the first 82 amino acids is lacking. The EpsK protein was suggested to play a role in the export of the exopolysaccharide by rendering the polymerized

exopolysaccharide more hydrophobic through a lipid modification. These data could suggest that the sequences in the region between Cps1G and Cps1H originated from *epsK*-like sequence. Hybridization experiments showed that this *epsK*-like region is also present in other serotype 1 strains as well as in serotype 14 strains (results not shown).

The function of most of the cloned serotype 9 genes can be established. Based on sequence similarity data the *cps9E* and *cps9F* genes could be glycosyltransferases (61, 24, 63, 64, 65). Moreover, the *cps9G* and *cps9H* genes showed similarity to genes located in regions involved in polysaccharide biosynthesis, but the function of these genes is unknown (68).

Cross-hybridization experiments using the individual *cps2*, *cps1* and *cps9* genes as probes showed that the *cps9G* and *cps9H* probes specifically hybridized with serotype 9 strains. Therefore, these are useful as tools for the identification of *S. suis* type 9 strains both for diagnostic purposes as well as in epidemiological and transmission studies. We previously developed a PCR method which can be used to detect *S. suis* strains in nasal and tonsil swabs of pigs (62). The method was for example used to identify pathogenic (EF-positive) strains of *S. suis* serotype 2 During the last years, beside *S. suis* type 2 strains, serotype 9 strains are frequently isolated from organs of diseased pigs. However, until now a rapid and sensitive diagnostic test was not available for type 9 strains. Therefore, the type 9 specific probes or the type 9 specific PCR is of great diagnostic value. The *cps1F*, *cps1G* and *cps1I* probes hybridized with serotype 1 as well as with serotype 14 strains. In coagglutination tests type 1 strains react with the anti-type 1 as well as with the anti-type 14 antisera (56). This suggests the presence of common epitopes between these serotypes. On the other hand type 1 strains agglutinated only with anti-type 1 serum (56,57), indicating that it is possible to detect differences between those serotypes.

The *cps2F*, *cps2G*, *cps2H*, *cps2I* and *cps2J* probes hybridized

with serotypes 2 and 1/2 only. Serotype 34 showed a weak hybridizing signal with the *cps2G* probe. As shown in agglutination tests type 1/2 strains react with sera directed against type 1 as well as with sera directed against type 2 strains (46). Therefore, type 1/2 shared antigens with both types 1 and 2. Based on the hybridization patterns of serotype 1/2 strains with the *cps1* and *cps2* specific genes, serotype 1/2 seemed to be more closely related to type 2 strains than to type 1 strains. In our current studies we identify type-specific genes, primers or probes which are used for the discrimination of serotypes 1, 14 and 2 and 1/2 and others of the 35 serotypes yet known. Furthermore, type-specific genes, primers or probes can now easily be developed for yet unknown serotypes, once they become isolated.

Cloning and characterization of a further part of the cps2 locus.

Based on the established sequence 11 genes, designated cps2L to cps2T, orf2U and orf2V , were identified. A gene homologous to genes involved in the polymerization of the repeating oligosaccharide unit (cps2O) as well as genes involved in the synthesis of sialic acid (cps2P to cps2T) were identified. Moreover, hybridization experiments showed that the genes involved in the sialic acid synthesis are present in S. suis serotype 1, 2, 14, 27 and 1/2. The "cps2M" and "cps2N" regions showed similarity to proteins involved in the polysaccharide biosynthesis of other gram-positive bacteria. However, these regions seemed to be truncated or were non-functional as the result of frame-shift or point mutations. At its 3'-end the cps2 locus contained two insertional elements ("orf2U" and "orf2V") both of which seemed to be non-functional.

To clone the remaining part of the cps2 locus, sequences of the 3'-end of pCPS26 (Fig. 1C) were used to identify a chromosomal fragment containing cps2 sequences located further downstream. This fragment was cloned in pKUN19 resulting in pCPS29. Using a similar approach we subsequently isolated the

plasmids pCPS30 and pCPS34 containing downstream cps2 sequences (Fig. 1C).

Analysis of the cps2 operon.

The complete nucleotide sequence of the cloned fragments was determined. Examination of the compiled sequence revealed the presence of : a sequence encoding the C-terminal part of Cps2K, six apparently functional genes (designated cps2O-cps2T) and the remnants of 5 different ancestral genes (designated "cps2L", "cps2M", "cps2N" , "orf2U" and "orf2V"). The latter genes seemed to be truncated or incomplete as the result of the presence of stop codons or frame-shift mutations (Fig. 1A). Neither potential promoter sequences nor potential stem-loop structures could be identified within the sequenced region. A ribosome-binding site precedes each ORF and the majority of the ORFs is very closely linked. Three intergenic gaps were found: one between "cps2M" and "cps2N" (176 nucleotides), one between cps2O and cps2P (525 nucleotides), and one between cps2T and "orf2U" (200 nucleotides). These and our above data show that Orf2X and Cps2A-Orf2T are part of a single operon.

A list of all loci and their properties is shown in Table 4. The "cps2L" region contained three potential ORFs, of 103, 79 and 152 amino acids, respectively, which were only separated from each other by stop codons. Only the first ORF is preceded by a potential ribosomal binding site and contained a methionine start codon. This suggests that "cps2L" originates from an ancestral cps2L gene, which coded for a protein of 339 amino acids. The function of this hypothetical Cps2L protein remains unclear so far: no significant homologies were found between Cps2L and proteins present in the data libraries. It is not clear whether the first ORF of the "cps2L" region is expressed into a protein of 103 amino acids. The"cps2M " region showed homology to the N-terminal 134 amino acids of the NeuA proteins of Streptococcus agalactiae and Escherichia coli (AB017355, 32). However,

although the "cps2 M" region contained a potential ribosome binding site, a methionine start codon was absent. Compared with the S. agalactiae sequence, the ATG start codon was replaced by a lysin encoding AAG codon. Moreover, the region homologous to the first 58 amino acids of the S. agalactiae NeuA (identity 77%) was separated from the region homologous to amino acids 59-134 of NeuA by a repeated DNA sequence of 100-bp (see later). In addition, the region homologous to amino acids 59 to 95 of NeuA (identity 32%) and the region homologous to the amino acids 96 to 134 of NeuA (identity 50%) were present in different reading frames. Therefore, the partial and truncated NeuA homologue is probably nonfunctional in S. suis. The "cps2N" region showed homology to CpsJ of S. agalactiae (accession no. AB017355). However, sequences homologous to the first 88 amino acids of CpsJ were lacking in S. suis. Moreover, the homologous region was present in two different reading frames. The protein encoded by the cps2O gene showed homology to proteins of several streptococci involved in the transport of the oligosaccharide repeating unit (accession no. AB017355), suggesting a similar function for Cps2O. The proteins encoded by the cps2P, cps2S and cps2T genes showed homology to the NeuB, NeuD and NeuA proteins of S. agalactiae and E. coli (accession no AB017355). Because the "cps2M" region also showed homology to NeuA of E. coli, the S. suis cps2 locus contains a functional neuA gene (cps2T) as well as a nonfunctional ("cps2M") gene. The mutual homology between these two regions showed an identity of 77% at the amino acid level over amino acids 1-58 and 49% over the amino acids 59-134. Cps2Q and Cps2R showed homology to the N-terminal and C-terminal parts of the NeuC protein of S. agalactiae and E. coli, respectively. This suggests that the function of the S. agalactiae NeuC protein in S. suis is likely fulfilled by two different proteins. In E. coli the neu genes are known to be involved in the synthesis of sialic acid. NeuNAc is synthesized from N-acetylmannosamine and phosphoenolpyruvate by NeuNAc synthetase. Subsequently, NeuNAc

is converted to CMP-NeuNAc by the enzyme CMP-NeuNAc synthetase. CMP-NeuNAc is the substrate for the synthesis of polysaccharide. In E. coli K1 NeuB is the NeuNAc synthetase, NeuA is the CMP-NeuNAc synthetase. NeuC has been implicated in the NeuNAc synthesis, but its precise role is not known. The precise role of NeuD is not known. A role of the Cps2P-Cps2T proteins in the synthesis of sialic acid can easily be envisaged, since the capsule of S. suis serotype 2 is rich in sialic acid. In S. agalactiae sialic acid has been shown to be critical to the virulence function of the type III capsule. Moreover, it has been suggested that the presence of sialic acid in capsule of bacteria which can cause meningitis may be important for the capacity of these bacteria to breach the blood-brain barrier. So far, however, the requirement of the sialic acid for virulence of S. suis remains unclear.

"Orf2U" and "Orf2V" showed homology to proteins located on two different insertional elements. "Orf2U" is homologous to IS1194 of Streptococcus thermophilus, whereas "Orf2V" showed homology to a putative transposase of Streptococcus pneumoniae. This putative transposase was recently found to be associated with the type 2 capsular locus of S. pneumoniae. Compared with the original insertional elements in S. thermophilus and S. pneumoniae, both "Orf2U" and "Orf2V" are likely to be non-functional due to frame shift mutations within their coding regions.

A striking observation was the presence of a sequence of 100 bp (Fig. 9) which was repeated three times within the cps2 operon. The sequence is highly conserved (between 94% and 98%) and was found in the intergenic regions between cps2G and cps2H, within "cps2M" and between cps2O and cps2P. No significant homologies were found between this 100-bp direct repeat sequence and sequences present in the data libraries, suggesting that the sequence is unique for S. suis.

Distribution of the cps2 sequences among the 35 S. suis serotypes. To examine the presence of sialic acid encoding genes in other S. suis serotypes, we performed cross-

hybridization experiments. DNA fragments of the individual cps2 genes were amplified by PCR, radiolabelled with 32P and hybridized to chromosomal DNA of the reference strains of the 35 different S. suis serotypes. As a positive control we used a probe specific for S. suis 16S rRNA. The 16S rRNA probe hybridized with almost equal intensities to all serotypes tested (Table 4). The "cps2L" sequence hybridized with DNA of serotype 1, 2, 14 and 1/2. The "cps2M", cps2O, cps2P, cps2Q, cps2R, cps2S and cps2T genes hybridized with DNA of serotype 1, 2, 14, 27 and 1/2. Because the cps2P-cps2T genes are most probably involved in the synthesis of sialic acid these results suggest that sialic acid is also a part of the capsule in the S. suis serotype 1, 2, 14, 27 and 1/2. This is in agreement with the finding that the serotypes 1, 2 and 1/2 possess a capsule that is rich in sialic acid. Although the chemical compositions of the capsules of serotype 14 and 27 are unknown, recent agglutination studies using sialic acid-binding lectins suggested the presence of sialic acid in S. suis serotype 14, but not in serotype 27. In these studies, sialic acid was also detected in serotypes 15 and 16. Since the latter observation is not in agreement with our hybridization studies, it might be that other genes, not homologous to the cps2P-cps2T genes, are responsible for the sialic acid synthesis in serotypes 15 and 16.

A probe based on"cps2N" sequences hybridized with DNA from serotypes 1, 2, 14 and 1/2. A probe specific for "orf2U" hybridized with serotypes 1, 2, 7, 14, 24, 27, 32, 34, and 1/2, whereas a probe specific for "orf2V" hybridized with many different serotypes. In addition, we prepared a probe specific for the 100-bp direct repeat sequence. This probe hybridized with the serotypes 1, 2, 13, 14, 22, 24, 27, 29, 32, 34 and 1/2 (Table 4). To analyze the number of copies of the direct repeat sequence within the S. suis serotype 2 chromosome, a Southern blot hybridization and analysis was performed. Therefore, chromosomal DNA of S. suis serotype 2 was digested with NcoI and hybridized with a 32P-labelled direct repeat

sequence. Only one hybridizing fragment, containing the three direct repeats present on the cps2 locus, was found (results not shown). This indicates that the 100-bp direct repeat sequence is only associated with the cps2 locus. In S. pneumoniae a 115-bp long repeated sequence was found to be associated with the capsular genes of serotypes 1, 3, 14 and 19F. In S. pneumoniae this 115-bp sequence was also found in the vicinity of other genes involved in pneumococcal virulence (hyaluronidase and neuraminidase genes). A regulatory role of the 115-bp sequence in co-ordinate control of these virulence-related genes was suggested.

To study the role of the capsule in resistance to phagocytosis and in virulence, we constructed two isogenic mutants in which capsule synthesis was disturbed. In 10cpsB, the *cps2B* gene was disturbed by the insertion of an antibiotic-resistance gene, whereas in 10cpsEF parts of the *cps2E* and *cps2F* genes were replaced. Both mutant strains seemed to be completely unencapsulated. Because the *cps 2* genes seemed to be part of an operon polar effects cannot be excluded. Therefore these data did not give any information about the role of Cps2B, Cps2E or Cps2F in the polysaccharide biosynthesis. However, the results clearly show that the capsular polysaccharide of *S. suis* type 2 is a surface component with antiphagocytic activity. *In vitro* wild type encapsulated bacteria are ingested by phagocytes at a very low frequency, whereas the mutant unencapsulated bacteria are efficiently ingested by porcine macrophages. Within 2 hours, over 99.6% of mutant bacteria were ingested and over 92% of the ingested bacteria were killed. Intracellularly, wild type as well as mutant strains seemed to be killed with the same efficiency. This suggests that the loss of capsular material is associated with loss of capacity to resist uptake by macrophages. This loss of resistance to *in vitro* phagocytosis was associated with a substantial attenuation of the virulence in germfree pigs. All pigs inoculated with the mutant strains survived the experiment and did not show any specific clinical

signs of disease. Only some aspecific clinical signs of disease could be observed. Moreover, mutant bacteria could be reisolated from the pigs. This supports the idea that, as in other pathogenic Streptococci, the capsule of *S. suis* acts as an important virulence factor. Transposon mutants prepared by Charland impaired in the capsule production showed a reduced virulence in pigs and mice. To construct these mutants the type 2 reference strain S735 was used. We previously showed that this strain is only weakly virulent for young pigs. Moreover, the insertion site of the transposon is unsolved sofar.

As a further example herein a rapid PCT test for *Streptococcus suis type* 7 is described.

Recent epidemiological studies on Streptococcus suis infections in pigs indicated that, besides serotypes 1, 2 and 9, serotype 7 is also frequently associated with diseased animals. For the latter serotype, however, no rapid and sensitive diagnostic methods are available. This hampers prevention and control programs. Here we describe the development of a type-specific PCR test for the rapid and sensitive detection of S. suis serotype 7. The test is based on DNA sequences of capsular (cps) genes specific for serotype 7. These sequences could be identified by cross-hybridization of several individual cps genes with the chromosomal DNAs of 35 different S. suis serotypes.

Streptococcus suis is an important cause of meningitis, septicemia, arthritis and sudden death in young pigs [69,70]. It can, however, also cause meningitis in man [71]. Attempts to control the disease are still hampered by the lack of sufficient knowledge about the epidemiology of the disease and the lack of effective vaccines and sensitive diagnostics.

S. suis strains can be identified and classified by their morphological, biochemical and serological characteristics [70, 73, 74]. Serological classification is based on the

presence of specific antigenic determinants. Isolated and biochemically characterized S. suis cells are agglutinated with a panel of specific sera. These typing methods are very laborious and time-consuming and can only be performed on isolated colonies. Moreover, it has been reported that nonspecific cross-reactions may occur among different types of S. suis [75, 76].

So far, 35 different serotypes have been described [7, 78, 79]. S. suis serotype 2 is the most prevalent type isolated from diseased pigs, followed by serotypes 9, and 1. However, recently serotype 7 strains were also frequently isolated from diseased pigs [80, 81, 82]. This suggests that infections with S. suis serotype 7 strains seemed to be an increasing problem. Moreover, the virulence of S. suis serotype 7 strains was confirmed by experimental infection of young pigs [83].

Recently, rapid and sensitive PCR assays specific for serotypes 2 (and 1/2), 1 (and 14) and 9 were developed [84]. These assays were based the cps loci of S. suis serotypes 2, 1 and 9 [84, 85]. However, until now no rapid and sensitive diagnostic test is available for S. suis serotype 7. Herein we describe the development of a PCR test for the rapid and sensitive detection of S. suis serotype 7 strains. The test is based on DNA sequences which form a part of the cps locus of S. suis serotype 7. Compared with the serological serotyping methods the PCR assay was a rapid, reliable and sensitive assay. Therefore, this test, in combination with the PCR tests which we previously developed for serotype 1, 2 and 9, will undoubtedly contribute to a more rapid and reliable diagnosis of S. suis and may facilitate control and eradication programs.

Materials and Methods

Bacterial strains, growth conditions and serotyping.

The bacterial strains and plasmids used in this study are listed in Table 7. The S. suis reference strains were obtained from M. Gottschalk, Canada. S. suis strains were grown in Todd-Hewitt broth (code CM189, Oxoid), and plated on Columbia agar blood base (code CM331, Oxoid) containing 6% (v/v) horse blood. E.coli strains were grown in Luria broth [86] and plated on Luria broth containing 1.5% (w/v) agar. If required, ampicillin was added to the plates. The S. suis strains were serotyped by the slide agglutination test with serotype-specific antibodies [70].

DNA techniques.

Routine DNA manipulations and PCR reactions were performed as described by Sambrook et al. [88]. Blotting and hybridization was performed as described previously [84,86].

DNA sequence analysis.

DNA sequences were determined on a 373A DNA Sequencing System (Applied Biosystems, Warrington, GB). Samples were prepared by use of a ABI/PRISM dye terminator cycle sequencing ready reaction kit (Applied Biosystems). Custom-made sequencing primers were purchased from Life Technologies. Sequencing data were assembled and analyzed using the McMollyTetra program. The BLAST program was used to search for protein sequences homologous to the deduced amino acid sequences.

PCR.

The primers used for the cps7H PCR correspond to the positions 3334-3354 and 3585-3565 in the S. suis cps7 locus. The sequences were:
5'-AGCTCTAACACGAAATAAGGC-3' and 5'-GTCAAACACCCTGGATAGCCG-3'. The reaction mixtures contained 10 mM Tris-HCl, pH 8.3; 1.5 mM

MgCl2; 50 mM KCl; 0.2 mM of each of the four deoxynucleotide triphosphates; 1 microM of each of the primers and 1U of AmpliTaq Gold DNA polymerase (Perkin Elmer Applied Biosystems, New Jersey). DNA amplification was carried out in a Perkin Elmer 9600 thermal cycler and the program consisted of an incubation for 10 min at 95oC and 30 cycles of 1 min at 95oC, 2 min at 56oC and 2 min at 72oC.

Results and discussion

Cloning of the seroytpe 7-specific cps genes.

To isolate the type-specific cps genes of S. suis serotype 7 we used the cps9E gene of serotype 9 as a probe to identify chromosomal DNA fragments of type 7 containing homologous DNA sequences [84]. A 1.6-kb PstI fragment was identified and cloned in pKUN19. This yielded pCPS7-1 (Fig. 11C). In turn, this fragment was used as a probe to identify an overlapping 2.7 kb ScaI-ClaI fragment. pGEM7 containing the latter fragment was designated pCPS7-2 (Fig. 11C).

Analysis of the cloned cps7 genes.

The complete nucleotide sequences of the inserts of pCPS7-1, pCPS7-2 were determined. Examination of the cps7 sequence revealed the presence of two complete and two incomplete open reading frames (ORFs) (Fig.11C). All ORFs are preceded by a ribosome-binding site. In accord with the data obtained for the cps1, cps2 and cps9 genes of serotypes 1, 2 and 9, respectively, the type 7 ORFs are very closely linked to each other. The only significant intergenic gap was that found between cps7E and cps7F (443 nucleotides). No obvious promoter sequences or potential stem-loop structures were found in this region. This suggests that, as in serotype 1, 2 and 9, the cps genes in serotype 7 form part of an operon.

An overview of the ORFs and their properties is shown in Table 8. As expected on the basis of the hybridization data [84], the Cps9E and Cps7E proteins showed a high similarity

(identity 99%, Table 8). Based on sequence comparisons between Cps9E and Cps7E, the PstI fragment of pCPS7-1 lacks the region encoding the first 371 codons of Cps7E. The C-terminal part of the protein encoded by the cps7F gene showed some similarity with the BplG protein of Bordetella pertussis [88], as well as with the C-terminal part of S. suis Cps2E [85]. Both BplG and Cps2E were suggested to have glycosyltransferase activity and are probably involved in the linkage of the first sugar to the lipid carrier [85,88]. The protein encoded by the cps7G gene showed similarity with the BlpF protein of Bordetella pertussis [88]. BplF is likely to be involved in the biosynthesis of an amino sugar, suggesting a similar function for Cps7G. The protein encoded by the cps7H gene showed similarity with the WbdN protein of E. coli [89] as well as with the N-terminal part of the Cps2K protein of S. suis [81]. Both WbdN and Cps2K were suggested to have glycosyltransferase activity [85, 89].

Serotype 7 specific cps genes.

To determine whether the cloned fragments in pCPS7-1 and pCPS7-2 contained serotype 7-specific DNA sequences, cross hybridization experiments were performed. DNA fragments of the individual cps7 genes were amplified by PCR, labelled with 32P, and used to probe spot blots of chromosomal DNA of the reference strains of 35 different S. suis serotypes. The results are summarized in Table 9. As expected, based on the data obtained with the cps9E probe [84], the cps7E probe hybridized with chromosomal DNA of many different S. suis serotypes. The cps7F and cps7G probes showed hybridization with chromosomal DNA of S. suis serotypes 4, 5, 7, 17, and 23. However, the cps7H probe hybridized with chromosomal DNA of serotype 7 only, indicating that this gene is specific for serotype 7.

Type specific PCR.

We tested whether we could use PCR instead of hybridization for the typing of the S. suis serotype 7 strains. For that purpose we selected an oligonucleotide primer set within the cps7H gene with which an amplified fragment of 251-bp was expected. In addition, we included in our analysis several S. suis serotype 7 strains, other than the reference strain. These strains were obtained from different countries and were isolated from different organs (Table 7). The results show that indeed a fragment of about 250-bp was amplified with all type 7 strains used (Fig. 12B), whereas no PCR products were obtained with serotype 1, 2 and 9 strains (Fig. 12A). This suggests that the PCR test, as described here, is a rapid diagnostic tool for the identification of S. suis serotype 7 strains. Until now such a diagnostic test was not available for serotype 7 strains. Together with the recently developed PCR assays for serotype 1, 2, 1/2, 14 and 9, this assay may be an important diagnostic tool to detect pigs carrying serotype 2, 1/2, 1, 14 ,9 and 7 strains and may facilitate control and eradication programs.

TABLE 1. Bacterial strains and plasmids

strain/plasmid	relevant characteristics	source/reference
Strain		
E.coli		
CC118	PhoA$^-$	(28)
XL2 blue	Stratagene	
E.coli		
XL2 blue	Stratagene	
S. suis		
10	virulent serotype 2 strain	(49)
3	serotype 2	(63)
17	serotype 2	(63)
735	reference strain serotype 2	(63)
T15	serotype 2	(63)
6555	reference strain serotype 1	(63)
6388	serotype 1	(63)
6290	serotype 1	(63)
5637	serotype 1	(63)
5673	serotype 1/2	(63)
5679	serotype 1/2	(63)
5928	serotype 1/2	(63)
5934	serotype 1/2	(63)
5209	reference strains serotype 1/2	(63)
5218	reference strain serotype 9	(63)
5973	serotype 9	(63)
6437	serotype 9	(63)
6207	serotype 9	(63)
reference strains	serotypes 1-34	(9, 56, 14)
S. suis		
10	virulent serotype 2 strain	(51)
10cpsB	isogenic cpsB mutant of strain 10	this work
10cpsEF	isogenic cpsEF mutant of strain 10	this work
Plasmid		
pKUN19	replication functions pUC, Amp^R	(23)
pGEM7Zf(+)	replication functions pUC, Amp^R	Promega Corp.
pIC19R	replication functions pUC, Amp^R	(29)
pIC20R	replication functions pUC, Amp^R	(29)
pIC-spc	pIC19R containing spc^R gene of pDL282	labcollection

pDL282	replication functions of pBR322 and pVT736-1, Amp^R, Spc^R	(43)
pPHOS2	pIC-spc containing the truncated *pho*A gene of pPHO7 as a *Pst*I-*Bam*HI fragment	this work
pPHO7	contains truncated *pho*A gene	(15)
pPHOS7	pPHOS2 containing chromosomal *S. suis* DNA	this work
pCPS6	pKUN19 containing 6 kb *Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS7	pKUN19 containing 3,5 kb *Eco*RI-*Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS11	pCPS7 in which 0.4 kb *Pst*I-*Bam*HI fragment of *cps*B gene is replaced by Spc^R gene of pIC-spc	this work (Fig.1)
pCPS17	pKUN19 containing 3.1 kb *Kpn*I fragment of *cps* operon	this work (Fig.1)
pCPS18	pKUN19 containing 1.8 kb *Sna*BI fragment of *cps* operon	this work (Fig.1)
pCPS20	pKUN19 containing 3.3 kb *Xba*I-*Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS23	pGEM7Zf(+) containing 1.5 kb *Mlu*I fragment of *cps* operon	this work (Fig.1)
pCPS25	pIC20R containing 2.5 kb *Kpn*I-*Sal*I fragment of pCPS17	this work (Fig.1)
pCPS26	pKUN19 containing 3.0 kb *Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS27	pCPS25 containing 2.3 kb *Xba*I (blunt)-*Cla*I fragment of pCPS20	this work (Fig.1)
pCPS28	pCPS27 containing the 1.2 kb *Pst*I-*Xho*I Spc^R gene of pIC-spc	this work (Fig.1)
pCPS29	pKUN19 containing 2.2 kb *Sac*I-*Pst*I fragment of cps operon	this work (Fig.1)
pCPS1-1	pKUN19 containing 5 kb *Eco*RV fragment of *cps* operon of type 1	this work (Fig.1)
pCPS1-2	pKUN19 containing 2.2 kb *Hin*dIII fragment of *cps* operon of type 1	this work (Fig.1)
pCPS9-1	pKUN19 containing 1 kb *Hin*dIII-*Xba*I fragment of *cps* operon of serotype 9	this work (Fig.1)
pCPS9-2	pKUN19 containing 4.0 kb *Xba*I-*Xba*I fragment of *cps* operon of serotype 9	this work (Fig.1)

Amp^R: ampicillin resistant
Spc^R: spectinomycin resistant
cps: capsular polysaccharide

Table 2. Properties of Orfs in the *cps* locus of *S. suis* serotype 2 and silimarities to gene product other bacteria

ORF	nucleotide position in sequence	number of amino acids	GC%	proposed function of gene product[1]	similar gene product (% identity)
Orf2Z	1 -719	240	44	Unknown	*B. subtilis* YitS (26%)
Orf2Y	2079-822	419	38	Transcription regulation	*B. subtilis* YcxD (39%)
Orf2X	2202-2934	244	39	Unknown	*H. influenzae* YAAA (24%)
Cps2A	3041-4484	481	39	Regulation	*S. pneumoniae* Cps19fA (58%)
Cps2B	4504-5191	229	40	Chain length determination	*S. pneumoniae* type 3 Orf1 (58%)
Cps2C	5203-5878	225	40	Chain length determination/ Export	*S. pneumoniae* Cps23fD (63%)
Cps2D	5919-6648	243	38	Unknown	*S. pneumoniae* CpsB (62%)
Cps2E	6675-8052	459	33	Glycosyltransferase	*S. pneumoniae* Cps14E (56%)
Cps2F	8089-9256	389	32	Glycosyl transferase	*S. pneumoniae* Cps23fT
Cps2G	9262-10417	385	36	Glycosyltransferase	*S. thermophilus* EpsF (25%)
Cps2H	10808-12176	457	31	Glycosyltransferase	*S. mutans* RGPEC,[N] (29%)
Cps2I	12213- 13443	410	29	CP polymerase	*S. pneumoniae* Cps23fI (48%)
Cps2J	13583-14579	332	29	Glycosyltransferase	*S. pneumoniae* Cps14J (31%)
Cps2K	14574-15576	334	37	Glycosyltransferase	*S.pneumoniae* Cps14J (40%)

Table 2 continued

"Cps2L"	15618-16635	103	37	Unknown	-
"Cps2M"	16811-17322	-	38	-	*S. agalactiae* CpsF^{N} (77%)
					E. coli NeuA ,^{N} (47%)
"Cps2N"	17559-18342	-	39	-	*S. agalactiae* CpsJ (43%)
Cps2O	18401-19802	476	40	Repeat unit transporter	*S. agalactiae* CpsK (41%)
Cps2P	20327-21341	338	39	Sialic acid synthesis	*S. agalactiae* NeuB (80%) *E. coli* NeuB (59%)
Cps2Q	21355-21865	170	42	Sialic acid synthesis	*S. agalactiae* NeuC^{N} (61%) *E. coli* NeuC^{N} (54%)
Cps2R	21933-22483	184	40	Sialic acid synthesis	*S. agalactiae* NeuC^{C} (55%) *E. coli* NeuC^{C} (40%)
Cps2S	22501-23125	208	42	Sialic acid synthesis	*E. coli* NeuD (32%)
Cps2T	23136-24366	395	40	CMP-NeuNAc synthetase	*S. agalactiae* CpsF (49%) *E.coli* NeuA (34%)
"Orf2U"	24566-25488	168	42	Transposase	*S. thermophilus* IS1194 (51%)
"Orf2V"	25691-26281	116	37	Transposase	*S. pneumoniae* orf1 (85%)

[1]Predicted by sequence similarity
[N] Similarity refers to the amino-terminal part of the gene product
[C] Similarity refers to the carboxy-terminal part of the gene product
ORFs between " " are truncated or non-functional as the result of frame-shift or point mutations

TABLE 3. Properties of ORFs in the *cps* genes of *S. suis* serotypes 1 and 9 and silimarities to gene products of other bacteria

ORF	nucleotide position in sequence	G + C%	number of amino acids	predicted mol. mass (kDa)	predicted pI	proposed function of gene product[1]	similar gene product (% identity)	reference/ accession nr.
Cps1E[2]	1-1363	34%	454	52.2	8.0	Glucosyltransferase	*Streptococcus suis* Cps2E (86%)	(26)
(48%)							*Streptococcus pneumoniae* Cps14E	(12)
Cps1F	1374-1821	33%	149	17.3	8.2	Unknown	*Streptococcus pneumoniae* Cps14F (83%)	(14)
Cps1G	1823-2315	25%	164	19.5	7.5	Glycosyltransferase	*Streptococcus pneumoniae* Cps14G(50%)	(14)
Cps1H	3035-4202	24%	389	45.5	8.4	CP polymerase	*Streptococcus pneumoniae* Cps14H (30%)	(14)
Cps1I	4197-					Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (38%)	(13)
							Lactoccocus lactis EpsG (31%)	(29)
							Streptococcus thermophilus EpsI (33%)	(28)
Cps1J						Glycosyltransferase	*Streptococcus pneumoniae* Cps14J ( )	

Table 3 continued

								(13)
Cps1K[3]		37%	278	32.5	7.8	Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (44%)	(13)
Cps9D[2]	1-646	37%	215	24.9	8.1	Unknown	*Streptococcus suis* Cps2D (89%)	(26)
Cps9E	680-					Glycosyltransferase	*Staphylococcus aureus* Cap1D (27%)	(18)
Cps9F		36%	200	22.3	8.2	Glycosyltransferase	*Staphylococcus aureus* Cap5M (52%)	(17)
Cps9G		35%	269	31.5	8.0	Unknown	*Actinobacillus actinomycetemcomitans* (43%)	(AB002668_4)
							Haemophilus influenzae Lsg (43%)	(O05081)
Cps9H[3]		30%	143	16.5	7.2	Unknown	*Yersinia enterolitica* RfbB (28%)	(33)

[1]Predicted by sequence similarity

[2] N-terminal part of protein is lacking

[3] C-terminal part of protein is lacking

Table 4. Hybridization of serotype 2 *cps* genes and neightbouring sequences with chromosomal DNA of other serotypes

serotypes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	1/2
DNA probes																																			
orf2Z	+	+	+	+	+	+	+	+	+	+	+	+	±	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	+	+	+	-	-	-	+
orf2Y	+	+	+	+	+	+	+	+	+	+	+	+	±	+	+	+	+	+	+	±	+	±	+	+	+	+	+	+	+	+	+	-	-	-	+
orf2X	+	+	+	+	+	+	+	+	+	+	+	+	±	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	+	+	+	-	-	-	+
cps2A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	+	+	+	-	-	-	+
cps2B	+	+	+	+	+	+	+	+	+	+	-	-	±	+	-	-	±	±	±	-	±	-	+	+	+	-	-	-	+	±	+	-	±	-	+
cps2C	+	+	+	+	+	+	+	+	+	+	+	-	±	+	-	±	-	-	-	-	-	-	+	+	+	-	+	±	-	-	+	-	±	-	+
cps2D	+	+	+	+	+	+	+	+	+	+	+	±	±	+	-	±	+	+	+	-	±	-	+	+	+	-	+	+	+	±	+	-	-	-	+
cps2E	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2F	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
cps2G	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	±	+
cps2H	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
cps2I	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
cps2J	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
cps2K	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
"cps2L"	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
"cps2M"	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
"cps2N"	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
cps2O	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2P	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2Q	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2R	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2S	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
cps2T	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+
"orf2U"	+	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	+	-	+	+
"orf2V"	+	+	±	±	±	-	±	-	-	-	-	-	-	+	+	-	+	+	±	-	-	±	+	-	-	+	-	-	-	-	+	-	±	+
100-bp repeat	+	+	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	-	-	+	-	+	+
16SrRNA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 5. Hybridization of serotypes 1 and 9 *cps* genes with chromosomal DNA of other *S. suis* serotypes

	DNA probes									
Serotype	*cps1E*	*cps1F*	*cps1G*	*cps1H*	*cps1I*	*cps9E*	*cps9F*	*cps9G*	*cps9H*	16rRNA
1	+	+	+	+	+	-	-	-	-	+
2	+	-	-	-	-	-	-	-	-	+
3	-	-	-	+	-	+	-	-	-	+
4	-	-	-	+	-	+	-	-	-	+
5	-	-	-	+	-	+	-	-	-	+
6	-	-	-	-	-	-	-	-	-	+
7	-	-	-	+	-	+	-	-	-	+
8	-	-	-	-	-	-	-	-	-	+
9	-	-	-	+	-	+	+	+	+	+
10	-	-	-	+	-	+	+	-	-	+
11	-	-	-	+	-	+	±	-	-	+
12	-	-	-	±	-	+	±	-	-	+
13	-	-	-	+	-	+	-	-	-	+
14	+	+	+	+	+	-	-	-	-	+
15	-	-	-	-	-	-	-	-	-	+
16	-	-	-	-	-	-	-	-	-	+
17	-	-	-	+	-	+	-	-	-	+
18	-	-	-	+	-	+	-	-	-	+
19	-	-	-	+	-	+	-	-	-	+
20	-	-	-	-	-	-	-	-	-	+
21	-	-	-	+	-	+	±	-	-	+
22	-	-	-	-	-	-	-	-	-	+

Table 5 continued

23	-	-	-	+	-	+	-	-	-	+
24	-	-	-	+	-	+	+	-	-	+
25	-	-	-	-	-	-	-	-	-	+
26	-	-	-	-	-	-	±	-	-	+
27	+	-	-	-	-	-	-	-	-	+
28	-	-	-	+	-	+	±	-	-	+
29	-	-	-	+	-	+	-	-	-	+
30	-	-	-	+	-	+	±	-	-	+
31	-	-	-	+	-	+	-	-	-	+
32	-	-	-	-	-	-	-	-	-	+
33	-	-	-	-	-	-	±	-	-	+
34	-	-	-	-	-	-	-	-	-	+
1/2	+	-	-	-	-	-	-	-	-	+

TABLE 6. Virulence of wild type and capsular mutant *S. suis* strains in germfree pigs

S. suis strains[1]	pigs/ group [n]	mortality[2] [%]	morbidity[3] [%]	clinical index of the group		fever index[7]	leuco-cyte index[8]	isolation of *S. suis* in pigs [n] per group in		
				spec symptoms[5]	non-spec. symptoms[6]			CNS	serosae	joints
10	4	100	100	11	88	43	44	2	3	4
10cpsB	4	0	0	0	10	1	3	1	3	2
10cpsEF	4	0	0	0	0	1	0	1	3	2

[1] strain10 in the wild type strain, strains 10cpsB and 10cpsEF are isogenic capsular mutant strains

[2] piglets which died spontaneously or had to be killed for animal welfare reasons

[3] only considering pigs with specific symptoms

[4] clinical index: % of observations which matched the described criteria

[5] specific symptoms: ataxia, lemeness on at lest one joint, stiffness

[6] non-specific symptoms: inappetance, depression

[7] % of observations in the experimental group with a body temperature > 40^0 C

[8] % of blood samples in the group in which number of granulocytes > 10^{10}/l

Tabl 7. Bacterial strains and plasmids

strain/plasmid	relevant characteristics
Strain	
E.coli	
XL2 blue	
S. suis	
reference strains	serotypes 1-34
5667	serotype 7, tonsil (1993)
7037	serotype 7, organs (1994)
7044	serotype 7, brains (1994)
7068	serotype 7 (1994)
7646	serotype 7 (1994)
7744	serotype 7, lungs (1996)
7759	serotype 7, joints (1996)
8169	serotype 7 (1997)
15913	serotype 7, meninges (1998)
Plasmid	
pKUN19replication functions pUC, Amp^R	
pGEM7Zf(+)	replication functions pUC, Amp^R
pCPS9-1	pKUN19 containing 1 kb *Hind*III-*Xba*I fragment of *cps* operon of serotype 9
pCPS9-2	pKUN19 containing 4.0 kb *Xba*I-*Xba*I fragment of *cps* operon of serotype 9
pCPS7-1	pKUN19 containing 1.6-kb *Pst*I fragment of *cps* operon of type 7
pCPS7-2	pGEM7 containing 2.7-kb *Sca*I-*Cla*I fragment of *cps* operon of type 7

*Amp^R: ampicillin resistant
cps: capsular polysaccharide

Table 8. Properties of Orfs in the *cps* genes of *S. suis* serotype 7 and similarities to gene products of other bacteria

Orf	nucleotide position in sequence	proposed function of gene product	similar gene product (% identity)
Cps7E	1-719	Glycosyltransferase	*Streptococcus suis* Cps9E (99%)
Cps7F	1164-1863	Glycosyltransferase	*Bordetella pertussis* BplG[1] (43%) *Streptococcus suis* Cps2E[1] (33%)
Cps7G	1872-3086	Biosynthesis amino sugar	*Bordetella pertussis* BplF (48%)
Cps7H	3104-3737	Glycosyltransferase	*Escherichia coli* WbdN (35%) *Streptococcus suis* Cps2K[2] (31%)

[1]similarity refers to the C-terminal part of the gene product
[2]similarity refers to the N-terminal part of the gene product

Table 9. Hybridization of serotype 7 *cps* probes with chromosomal DNA of *S. suis* serotypes

serotypes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	1/2
DNA probes																																			
cps7E	-	-	+	+	+	-	+	-	+	+	+	+	+	-	-	-	+	+	+	-	+	-	+	+	-	-	-	-	+	+	+	-	-	-	-
cps7F	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
cps7G	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
cps7H	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16SrRNA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

LEGENDS TO FIGURES

Figure 1.

Organization of the cps2 gene cluster of S. suis type 2.

(A) Genetic map of the cps2 gene cluster. The shadowed arrows represent potential ORFs. Interrupted ORFs indicate the presence of stop codons or frame-shift mutations. Gene designations are indicated below the ORFs. The closed arrows indicate the position of the potential promoter sequences. | indicates the position of the potential transcription regulator sequence. ||| indicates the position of the 100-bp repeated sequence.

(B) Physical map of the cps2 locus.

Restriction sites are as follows: A: AluI; C: ClaI; E, EcoRI; H, HindIII; K, KpnI; M, MluI; N, NsiI; P, PstI; S, SnaBI; Sa: SacI; X, XbaI.

(C) The DNA fragments cloned in the various plasmids.

Figure 2

Ethidium bromide stained agarose gel showing PCR products obtained with chromosomal DNA of *S.suis* strains belonging to the serotypes 1,2, ½, 9 and 14 and *cps2J, cps1I* and *cps9H* primer sets as described in Materials and Methods. (A) *cps1I* primers.

(B) *cps2J* primers and (C) *cps9H* primers. Lanes 1-3: serotype 1 strains; lanes 4-6: serotype 2 strains; lanes 7-9: serotype ½ strains; lanes 10-12: serotype 9 strains and lanes 13-15: serotype 14 strains.

(B) Ethidium bromide stained agarose gel showing PCR products obtained with tonsillar swabs collected from pigs carrying *S.suis* type 2, type 1 or type 9 strains and *cps2j, cps1I* and *cpsH* primer sets as described in Materials and Methods. Bacterial DNA suitable for PCR was prepared by using the multiscreen methods as described previously (20). (A) *cps1I* primers. (B) *cps2J* primers and (C) *cps9H* primers. Lanes 1-3: PCR products obtained with tonsillar swabs collected from pigs

carrying *S.suis* type 1 strains; lanes 4-6: PCR products obtained with tonsillar swabs collected from pigs carrying *S.suis* type 2 strains; lanes 7-9: PCR procucts obtained with tonsillar swabs collected from pigs carrying *S.suis* type 9 strains; lanes 10-12: PCR products obtained with chromosomal DNA from serotype 9, 2 and 1 strains respectively; lane 13: negative control, no DNA present.

Figure 3
CPS2 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 4
CPS1 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 5
CPS9 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 6
CPS7 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 7
Alignments of the N-terminal parts of Cps2J and Cps2K. Identical amino acids are marked by bars. The amino acids shown in bold are also conserved in Cps14I, Cps14J of *S. pneumoniae* and several other glycosyltransferases (19). The aspartate residues marked by asterics are strongly conserved.

Figure 8
Transmission electron micrographs of thin sections of various *S. suis* strains.
(A) wild type strain 10;

(B) mutant strain 10cpsB;

(C) mutant strain 10cpsEF.

Bar = 100 nm

Figure 9

(A) Kinetics of phagocytosis of wild type and mutant *S. suis* strains by porcine alveolair macrophages. Phagocytosis was determined as described in Materials and Methods. The Y-axis represents the number of CFU per milliliter in the supernatant fluids as determined by plate counting, the X-axis represents time in minutes.

□ wild type strain 10;

o mutant strain 10cpsB;

Δ mutant strain 10cpsEF.

(B) Kinetics of intracellular killing of wild type and mutant *S. suis* strains by porcine AM. The intracellular killing was determined as described in Material and Methods. The Y-axis represents the number of CFU per ml in the supernatant fluids after lysis of the macrophages as determined by plate counting, the X-axis represents time in minutes.

□ wild type strain 10;

o mutant strain 10cpsB;

Δ mutant strain 10cpsEF.

Figure 10

Nucleotide sequence alignment of the highly conserved 100-bp repeated element.

1) 100-bp repeat between cps2G and cps2H

2) 100-bp repeat within "cps2M"

3) 100-bp repeat between cps2O and cps2P

Figure 11. The cps2, cps9 and cps7 gene clusters of S. suis serotypes 2, 9 and 7.

(A) Genetic organization of the cps2 gene cluster [84]. The large arrows represent potential ORFs. Gene designations are indicated below the ORFs. Identically filled arrows represent ORFs which showed homology. The small closed arrows indicate the position of the potential promoter sequences. | indicates the position of the potential transcription regulator sequence.

(B) Physical map and genetic organization of the cps9 gene cluster [15]. Restriction sites are as follows: B: BamHI; P: PstI; H: HindIII; X:XbaI. The DNA fragments cloned in the various plasmids are indicated. The open arrows represent potential ORFs.

(C) Physical map and genetic organization of the cps7gene cluster. Restriction sites are as follows: C: ClaI; P: PstI; Sc: ScaI. The DNA fragments cloned in the various plasmids are indicated. The open arrows represent potential ORFs.

Figure 12 (A) Ethidium bromide stained agarose gel showing PCR products obtained with chromosomal DNA of S. suis strains belonging to the serotypes 1, 2, 9 and 7 and the cps7H primer set. Strain designations are indicated above the lanes. C: negative control, no DNA present. M: molecular size marker (lambda digested with EcoRI and HindIII).

(B) Ethidium bromide stained agarose gel showing PCR products obtained with serotype 7 strains collected in different countries and from different organs. Bacterial DNA suitable for PCR was prepared by using the multiscreen method as described previously [89]. Strain designations are indicated above the lanes. M: molecular size marker (lambda digested with EcoRI and HindIII).

REFERENCES

1. Arends, J. P., and H. C. Zanen. 1988. Meningitis caused by *Streptococcus suis* in humans. Rev. Infect. Dis. 10:131-137.

2. Arrecubieta, C., E. Garcia, and R. Lopez. 1995. Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. Gene 167: 1-7

3. Arrecubieta, C., R. Lopez, and E. Garcia. 1994. Molecular characterization of *cap3A*, a gene from the operon required for the synthesis of the capsule of *Streptococcus pneumoniae* type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. J. Bacteriol. 176: 6375-6383.

4. Clifton-Hadley, F.A. 1983. *Streptococcus suis* type 2 infections. Br. Vet. J. 139:1-5.

5. Charland, N., J. Harel, M. Kobisch, S. Lacasse, and M. Gottschalk. 1998. Streptococcus suis serotype 2 mutants deficient in capsular expression. Microbiol. 144:325-332.

6. Cross, A. S. 1990. The biological significance of bacterial encapsulation. Curr. Top. Microbiol. Immunol. 150: 87-95.

7. Elliott, S. D. and J. Y. Tai . 1978. The type specific polysaccharide of *Streptococcus suis*. J. Exp. Med. 148: 1699-1704.

8. Feder, I., M. M. Chengappa, B. Fenwick, M. Rider and J. Staats. 1994. Partial characterization of *Streptococcus suis* type 2 hemolysin. J. Clin. Microbiol. 32:1256-1260.

9. Gottschalk, M., R. Higgins, M. Jacques, M. Beaudoin, and J. Henrichsen. 1991. Characterization of six new capsular types (23 through 28) of *Streptococcus suis*. J. Clin. Microbiol. 29:2590-2594.

10. Gottschalk, M., S. Lacouture, and J. D. Dubreuil. 1995. Characterization of *Streptococcus suis* type 2 haemolysin.

Microbiology 141:189-195.

11. Gottschalk, M., A. Lebrun, M. Jacques, and R. Higgins. 1990. Haemagglutination properties of *Streptococcus suis*. J. Clin. Microbiol. 28: 2156-2158.

12. Guidolin, A., J. M. Morona, R. Morona, D. Hansman, and J. C. Paton. 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. 1994. Infect. Immun. 62: 5384-5396.

13. Guitierrez, C., and J. C. Devedjian. 1989. Plasmid facilitating *in vitro* construction of PhoA fusions in *Escherichia coli*. Nucl. Acid. Res. 17: 3999.

14. Higgins, R., M. Gottschalk, M. Boudreau, A. Lebrun, and J. Henrichsen. 1995. Description of six new capsular types (28 through 34) of *Streptococcus suis*. J. Vet. Diagn. Invest. 7:405-406

15. Jacobs, A. A., P. L. W. Loeffen, A. J. G. van den Berg, and P. K. Storm. 1994. Identification, purification and characterization of a thiol-activated hemolysin (suilysin) of *Streptococcus suis*. Infect. Immun. 62: 1742-1748.

16. Jacques, M., M. Gottschalk, B. Foiry B.and R. Higgins. 1990. Ultrastructural study of surface components of *Streptococcus suis*. J. Bacteriol. 172:2833-2838.

17. Klein P., M. Kanehisa and C. DeLisi. 1985. The detection and classification of membrane spanning proteins. Biochim. Biophys. Acta. 851: 468-476.

18. Kolkman, M. A. B., D. A. Morrison, B. A. M. van der Zeijst, and P. J. M. Nuijten. 1996. The capsule polysaccharide synthesis locus of *Streptococcus pneumoniae* serotype 14: identification of the glycosyl transferase gene *cps14E*. J. Bacteriol. 178: 3736-3541.

19. Kolkman, M. A. B., W. Wakarchuk, P. J. M. Nuijten, and B. A. M. van der Zeijst. 1997. Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the

tetrasaccharide subunit. Mol. Microbiol. 26: 197-208.

20. Kolkman, M. A. B., B. A. M. van der Zeijst and P. J. M. Nuijten. 1997. Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14. J. Biol. Chem. 272: 1950219508.

21. Konings, R. N. H., E. J. M. Verhoeven, and B. P H. Peeters. 1987. pKUN vectors for the separate production of both DNA strands of recombinant plasmids. Methods Enzymol. 153: 12-34.

22. Korolik, V., B. N. Fry, M. R. Alderton, B. A. M. van der Zeijst, and P. J. Coloe. 1997. Expression of *Campylobacter hyoilei* lipo-oligosaccharide (LOS) antigens in *Escherichia coli*. Microbiol. 143: 3481-3489.

23. Leij, P. C. J., R. van Furth, and T. L. van Zwet. 1986. In vitro determination of phagocytosis and intracellular killing of polymorphonuclear and mononuclear phagocytes. In Handbook of Experimental Immunology, vol. 2. Cellular Immunology, pp. 46.1-46.21. Edited by D. M. Weir, L. A. Herzenberg, C. Blackwell and L. A. Herzenberg. Blackwell Scientific Publications, Oxford.

24. Lin, W. S., T. Cunneen, and C. Y. Lee. 1994. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. J. Bacteriol. 176: 7005-7016.

25. Liu, D., A. M. Haase, L. Lindqvist, A.A. Lindberg, and P. R. Reeves. 1993. Glycosyl transferases of O-antigen biosynthesis in *Salmonella enteritica*: Identification and characterization of transferase genes of group B, C2, and E1. J. Bacteriol. 175: 3408-3413.

26. Manoil, C., and J. Beckwith. 1985. A transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA 82: 8129-8133.

27. Marsh, J. L., M. Erfle, and E. J. Wykes. 1984. The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene

32:481-485.

28. Miller, J. 1972. Experiments in Molecular Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

29. Morona, J. K., R. Morona, and J. C. Paton. 1997. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthesis pathway. Mol. Microbiol. 23: 761-763.

30. Muñoz, R., M. Mollerach, R. López and E. Garcia. 1997. Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae;* formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. Mol. Microbiol. 25: 79-92.

31. Pearce B. J., Y. B. Yin, and H. R. Masure. 1993. Genetic identification of exported proteins in *Streptococcus pneumoniae*. Mol. Microbiol. 9: 1037-1050.

32. Roberts, I. S. 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Ann. Rev. Microbiol. 50: 285-315.

33. Rossbach, S., D. A. Kulpa, U. Rossbach, and F. J. de Bruin. 1994. Molecular and genetic characterization of the rhizopine catabolism (mocABRC) genes of *Rhizobium meliloti* L5-30. Mol. Gen. Genet. 245: 11-24.

34. Rubens, C. E., L. M. Heggen, R. F. Haft, and R. M. Wessels. 1993. Identification of *cpsD*, a gene essential for type III capsule expression in group B streptococci. Mol. Microbiol. 8: 843-855.

35. Rubens, C. E., L. M. R. Wessels, L. M. Heggen, and D. L. Kasper. 1987. Transposon mutagenesis of type III group B Streptococcus: correlation of capsule expression with virulence. Proc. Natl. Acad. Sci. USA 84:7208-7212.

36. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Second edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. New York.

37. Smith, H. E., U. Vecht, H. J. Wisselink, N. Stockhofe-Zurwieden, Y. Biermann, and M. A. Smits. 1996. Mutants of

Streptococcus suis types 1 and 2 impaired in expression of muramidase-released protein and extracellular protein induce disease in newborn germfree pigs. Infect Immun. 64: 4409-4412.

38. Smith, H. E., H. J. Wisselink, U. Vecht, A. L. J. Gielkens and M. A. Smits. 1995. High-efficiency transformation and gene inactivation in *Streptococcus suis* type 2. Microbiol. 141: 181-188.

39. Sreenivasan, P. K., D. L. LeBlanc, L. N. Lee, and P. Fives-Taylor. 1991. Transformation of *Actinobacillus actinomycetemcomitans* by electroporation, utilizing constructed shuttle plasmids. Infect. Immun. 59: 4621-4627.

40. Stringele F., J.-R. Neeser, and B. Mollet. 1996. Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. J. Bacteriol. 178: 1680-1690.

41. Stockhofe-Zurwieden, N., U. Vecht, H. J. Wisselink, H.van Lieshout, and H. E. Smith. 1996. Comparative studies on the pathogenicity of different *Streptococcus suis* serotype 1 strains. In Proceedings of the 14th IPVS Congress. pp. 299.

42. van Kranenburg, R., J. D. Marugg, I. I. van Swam, N. J. Willem and W. M. de Vos. 1997. Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis* Mol. Microbiol. 24: 387-397.

43. van Leengoed, L. A., E. M. Kamp, and J. M. A. Pol. 1989. Toxicity of *Haemophilus pleuropneumoniae* to porcine lung macrophages. Vet. Microbiol. 19: 337-349.

44. van Leengoed, L. A. M. G., U. Vecht, and E. R. M. Verheyen. 1987. *Streptococcus suis* type 2 infections in pigs in The Netherlands (part two). Vet Quart. 9, 111-117.

45. Vecht, U., J. P. Arends, E. J. van der Molen, and L. A. M. G. van Leengoed. 1989. Differences in virulence between two strains of *Streptococcus suis* type 2 after experimentally induced infection of newborn germfree pigs. Am. J. Vet. Res. 50:1037-1043.

46. Vecht, U., L. A. M. G. van Leengoed, and E. R. M.

Verheyen. 1985. *Streptococcus suis* infections in pigs in The Netherlands (part one). Vet. Quart. 7:315-321

47. Vecht, U., H. J. Wisselink, M. L. Jellema, and H. E. Smith. 1991. Identification of two proteins associated with virulence of *Streptococcus suis* type 2. Infect. Immun. 59:3156-3162.

48. Vecht, U., H. J. Wisselink, N. Stockhofe-Zurwieden, and H. E. Smith. 1996. Characterization of virulence of the Streptococcus suis serotype 2 reference strain Henrichsen S 735 in newborn gnotobiotic pigs. Vet. Microbiol. 51:125-136.

49. Vecht, U., H. J. Wisselink, J. E. van Dijk, and H. E. Smith. 1992. Virulence of *Streptococcus suis* type 2 strains in newborn germfree pigs depends on phenotype. Infect. Immun. 60:550-556.

50. Wagenaar, F., G. L. Kok, J. M. Broekhuijsen-Davies, and J. M. A. Pol. 1993. Rapid cold fixation of tissue samples by microwave irradiation for use in electron microscopy. Histochemical J. 25: 719-725.

51. Wessels, M. R.and M. S. Bronze. 1994. Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. Proc. Natl. Acad. Sci. USA 91: 12238-12242.

52. Wessels, M. R., A. E. Moses, J. B. Goldberg, and T. J. DiCesare. 1991. Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. Proc. Natl. Acad. Sci. USA. 88: 8317-8321.

53. Yamane, K., M. Kumamano, and K.Kurita. 1996. The 25°-36° region of the *Bacillus subtilis* chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes. Microbiol. 142: 3047-3056.

54. Butler, J. C., R. F. Breiman, H. B. Lipman, J. Hofmann, and R. R. Facklam. 1995. Serotype distribution of *Streptococcus pneumoniae* infections among preschool children in the United States, 1978-1994: implications fro development of a conjugate vaccine. J. Infect. Dis. 171: 885-889.

55. Charland, N., M. Jacques, S. Lacoutre and M. Gottschalk.

1997. Characterization and protective activity of a monoclonal antibody against a capsular epitope shared by *Streptococcus suis* serotypes 1, 2 and 1/2. Microbiol. 143:3607-3614.

56. Gottschalk, M., R. Higgins, M. Jacques, K. R. Mittal, and J. Henrichsen. Description of 14 new capsular types of *Streptococcus suis*. J. Clin. Microbiol. 27:2633-2636.

57. Heath, P. J., B. W. Hunt, and J. P. Duff. 1996. *Streptococcus suis* serotype 14 as a cause of pig disease in the UK. Vet. Rec. 2:450-451.

58. Hommez, J., L. A. Devrieze, J. Henrichsen, and F. Castryck. 1986. Identification and characterization of *Streptococcus suis*. Vet. Microbiol. 16:349-355.

59. Killper-Balz, R., and K. H. Schleifer. 1987. *Streptococcus suis* sp. nov. nom. rev. Int. J. Syst. Bacteriol. 37:160-162.

60. Kolkman, M. A. B., B. A. M. van der Zeijst, and P. J. M. Nuijten. 1998. Diversity of capsular polysaccharide synthesis gene clusters in Streptococcus pneumoniae. Submitted for publication.

61. Lee, J. C., S. Xu, A. Albus, and P. J. Livolsi. 1994. Genetic analysis of type 5 capsular polysaccharide expression by *Staphylococcus aureus*. J. Bacteriol. 176:4883-4889.

62. Reek, F. H., M. A. Smits, E. M. Kamp, and H. E. Smith. 1995. Use of multiscreen plates for the preparation of bacterial DNA suitable for PCR. BioTechniques 19: 282-285.

63. Sau, S., N. Bhasin, E. R. Wann, J. C. Lee, T. J. Foster, and C. Y. Lee. 1997. The *Staphylococcus aureus* allelic genetic loci for serotype 5 and 8 capsule expression contain the type-specific genes flanked by common genes. Microbiol. 143: 2395-2405.

64. Sau, S., and C. Y. Lee. 1996. Cloning of type 8 capsule genes and analysis of gene clusters for the production of different capsular polysaccharides in *Staphylococcus aureus*. J. Bacteriol. 178: 2118-2126.

65. Sau, S., and C. Y. Lee. 1997. Molecular characterization and transcriptional analysis of type 8 capsule genes in

Staphylococcus aureus. J. Bacteriol. 179:1614-1621.

66. Smith, H. E., M. Rijnsburger, N. Stockhofe-Zurwieden, H. J. Wisselink, U. Vecht, and M. A. Smits. 1997. Virulent strains of *Streptococcus suis* serotype 2 and highly virulent strains of *Streptococcus suis* serotype 1 can be recognized by a unique ribotype profile. J. Clin. Microbiol. 35:1049-1053.

67. Yamazaki, M., L. Thorne, M. Mikolajczak, R. W. Armentrout, and T. J. Pollock. 1996. Linkage of genes essential for synthesis of a polysaccharide capsule in *Sphingomonas* strain S88. J. Bacteriol. 178:2676-2687.

68. Zhang, L., A.Al-Hendy, P. Toivanen. and M. Skurnik. 1993. Genetic organization and sequence of the *rfb* gene cluster of *Yersinia enterolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. Mol. Microbiol. 9:309-321.

69 Clifton-Hadley, F.A. (1983). Streptococcus suis type 2 infections. Br. Vet. J. 139, 1-5.

70 Vecht, U., van Leengoed, L. A. M. G. and Verheyen, E. R. M. (1985). Streptococcus suis infections in pigs in The Netherlands (part one). Vet. Quart. 7, 315-321

71 Arends, J. P. and Zanen, H. C.(1988). Meningitis caused by Streptococcus suis in humans. Rev. Infect. Dis. 10, 131-137.

72 Hommez, J., Devrieze, L.A., Henrichsen, J. and Castryck, F. (1986). Identification and characterization of Streptococcus suis. Vet. Microbiol. 16, 349-355.

73 Killper-Balz, R. and Schleifer, K. H. (1987). Streptococcus suis sp. nov. nom.rev. Int. J. Syst. Bacteriol. 37, 160-162.

74 Gottschalk, M., Higgins, R. and Jacques, M. (1993). Production of capsular material by Streptococcus suis serotype 2 under different conditions. Can. J. Vet. Res. 57, 49-52.

75 Higgins, R. and Gottschalk, M. (1990). Un update on Streptococcus suis identification. J. Vet. Diagn. Invest. 2, 249-252.

76 Gottschalk, M., Higgins, R., Jacques, M., Beaudoin, M. and Henrichsen, J. (1991). Characterization of six new capsular types (23 through 28) of Streptococcus suis. J. Clin. Microbiol. 29, 2590-2594.

77 Gottschalk, M., Higgins, R., Jacques, M., Mittal, K. R. and Henrichsen, J. (1989) Description of 14 new capsular types of Streptococcus suis. J. Clin. Microbiol. 27, 2633-2636.

78 Higgins, R., Gottschalk, M., Boudreau, M.,Lebrun, A.and Henrichsen, J. (1995). Description of six new capsular types (28 through 34) of Streptococcus suis. J. Vet. Diagn. Invest. 7, 405-406

79 Aarestrup, F. M., Jorsal, S. E. and Jensen, N. E. (1998). Serological characterization and antimicrobial susceptibility of Streptococcus suis isolates from diagnostic samples in Denmark during 1995 and 1996. Vet. Microbiol. 15, 59-66.

80 MacLennan, M., Foster, G., Dick, K., Smith, W. J. and Nielsen, B. (1996). Streptococcus suis serotypes 7, 8 and 14 from diseased pigs in Scotland. Vet Rec. 139, 423-424.

81 Sihvonen, L., Kurl, D. N. and Henrichsen, J. (1988). Streptococcus suis isolates from pigs in Finland. Acta Vet. Scand. 29, 9-13.

82 Boetner, A. G., Binder, M. and Bille-Hansen, V. (1987). Streptococcus suis infections in Danish pigs and experimental infection with Streptococcus suis serotype 7. Acta Path. Microbiol. Immunol. Scand. Sect. B, 95, 233-239.

83 Smith, H. E., Veenbergen, V., van der Velde, J., Damman, M., Wisselink, H. J. and Smits, M. A. (1999). The cps genes of Streptococcus suis serotypes 1, 2 and 9: development of rapid serotype-specific PCR assays. J. Clin. Microbiol. submitted

84 Smith, H. E., Damman, M., van der Velde, J., Wagenaar, F., Wisselink, H. J., Stockhofe-Zurwieden, N. and Smits, M. A. (1999). Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor. Infect. Immun. 67, 1750-1756.

85 Miller, J. (1972). Experiments in Molecular Genetics.

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

86 Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

87 Allen, A. and Maskell, D. (1996). The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in Bordetella pertussis. Mol. Microbiol. 19, 37-52.

88 Wang, L. and Reeves, P. R. (1998). Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes. Infect. Immun. 66, 3545-3551.

89 Wisselink, H. J., Reek, F. H., Vecht, U., Stockhofe-Zurwieden, N., Smits, M. A. and Smith, H. E. (1999). Detection of virulent strains of Streptococcus suis type 2 and highly virulent strains of Streptococcus suis type 1 in tonsillar specimens of pigs by PCR. Vet. Microbiol. 67, 143-157.

90 Konings, R. N. H., Verhoeven, E. J. M. and Peeters, B. P. H. (1987). pKUN vectors for the separate production of both DNA strands of recombinant plasmids. Methods Enzymol. 153, 12-34.

CLAIMS

1. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derived thereof.

2. A nucleic acid according to claim 1 encoding a *Streptococcus suis* serotype-specific central region, preferably encoding at least one enzyme or fragment thereof involved in polysaccharide biosynthesis.

3. A nucleic acid according to claim 1 or 2 hybridising to a nucleic acid encoding a gene derived from a *Streptococcus suis* serotype 1, 2 or 9 capsular gene cluster.

4. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 2 or a gene or gene fragment derived thereof, preferably as identified in Figure 3.

5. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 1 or a gene or gene fragment derived thereof, preferably as identified in Figure 4.

6. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 9 or a gene or gene fragment derived thereof, preferably as identified in Figure 5.

7. A nucleic acid probe or primer derived from a nucleic acid according to anyone of claims 1 to 6 allowing species or serotype specific detection of *Streptococcus suis*.

8. A probe or primer according to claim 7 provided with at least one reporter molecule.

9. A diagnostic test comprising a probe or primer according to claim 7 or 8.

10. A protein or fragment thereof encoded by a nucleic acid according to anyone of claims 1 to 6.

11. A protein or fragment according to claim 10 capable of polysaccharide biosynthesis.

12. A method to produce a *Streptococcus suis* capsular antigen comprising using a protein or fragment according to claim 11.

13. A *Streptococcus suis* capsular antigen obtainable by a method according to claim 12.

14. A vaccine comprising an antigen according to claim 13 and further comprising a suitable carrier or adjuvant.

15. A recombinant *Streptococcus suis* mutant provided with a modified capsular gene cluster.

16. A recombinant micro-organism comprising at least a part of a capsular gene cluster of *Streptococcus suis*.

17. A recombinant micro-organism according to claim 16 comprising a lactic acid bacterium.

18. A vaccine comprising a mutant according to claim 15 or a micro-organism according to claim 16 or 17.

19. A vaccine according to claim 18 comprising a *Streptococcus* mutant deficient in capsular expression.

20. A vaccine according to claim 19 wherein said *Streptococcus* mutant has been derived by recombinant techniques, preferably through homologous recombination.

21. A vaccine according to claim 19 or 20 wherein said mutant is capable of surviving in an immune-competent host.

22. A vaccine according to claim 21 wherein said mutant is capable of surviving at least 4-5 days, preferably at least 8-10 days, in said host.

23. A vaccine according to any of claims 19 to 22 comprising a mutant capable of expressing a *Streptococcus* virulence factor or antigenic determinant.

24. A vaccine according to any of claims 19 to 23 comprising a mutant capable of expressing a non-*Streptococcus* protein.

25. A vaccine according to claim 24 wherein said non-Streptococcus protein has been derived from a pathogen.

26. A method for controlling or eradicating a Streptococcal disease in a population comprising vaccinating subjects in said population with a vaccine according to anyone of claims 18 to 25.

27. A method for controlling or eradicating a Streptococcal

disease comprising testing a sample collected from at least one subject in a population partly or wholly vaccinated with a vaccine according to anyone of claims 19 to 25 for the presence of encapsulated Streptococcal strains.

28. A method for controlling or eradicating a Streptococcal disease comprising testing a sample collected from at least one subject in a population partly or wholly vaccinated with a vaccine according to anyone of claims 19 to 25 for the presence of capsule-specific antibodies directed against Streptococcal strains.

29. A method for controlling or eradicating a Streptococcal disease in a population comprising selecting subjects in said population vaccinated with a vaccine according to anyone of claims 19 to 25 and testing a sample collected from at least one subject in said population for the presence of encapsulated Streptococcal strains and/or for the presence of capsule-specific antibodies directed against Streptococcal strains.

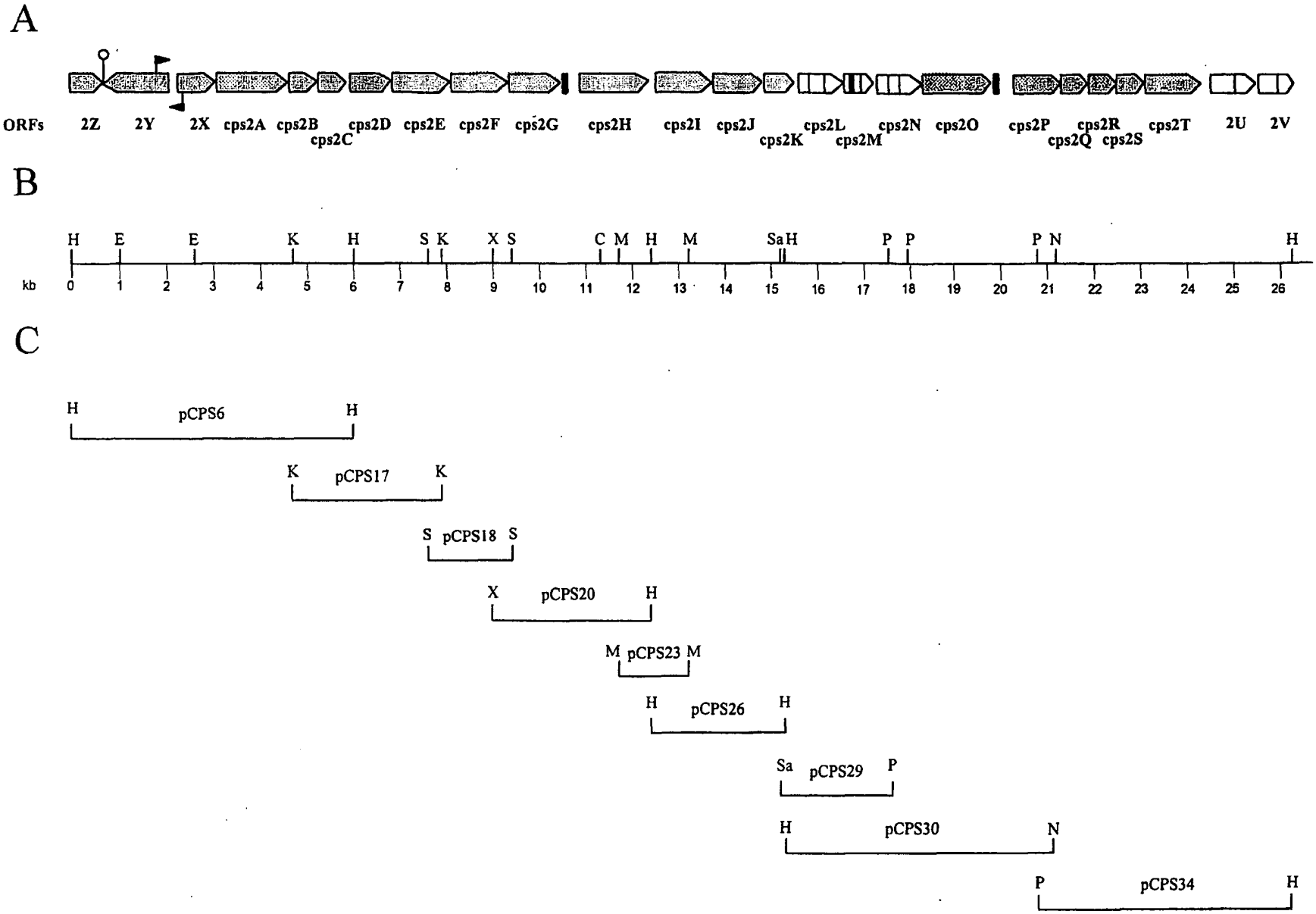
A
ORFs
2Z
2Y
2X
cps2A
cps2B
cps2C
cps2D
cps2E
cps2F
cps2G
cps2H
cps2I
cps2J
cps2K
cps2L
cps2M
cps2N
cps2O
cps2P
cps2Q
cps2R
cps2S
cps2T
2U
2V
B
kb
H E E K H S K X S C M H M Sa H P P P N H
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
C
H pCPS6 H
K pCPS17 K
S pCPS18 S
X pCPS20 H
M pCPS23 M
H pCPS26 H
Sa pCPS29 P
H pCPS30 N
P pCPS34 H

Fig. 1

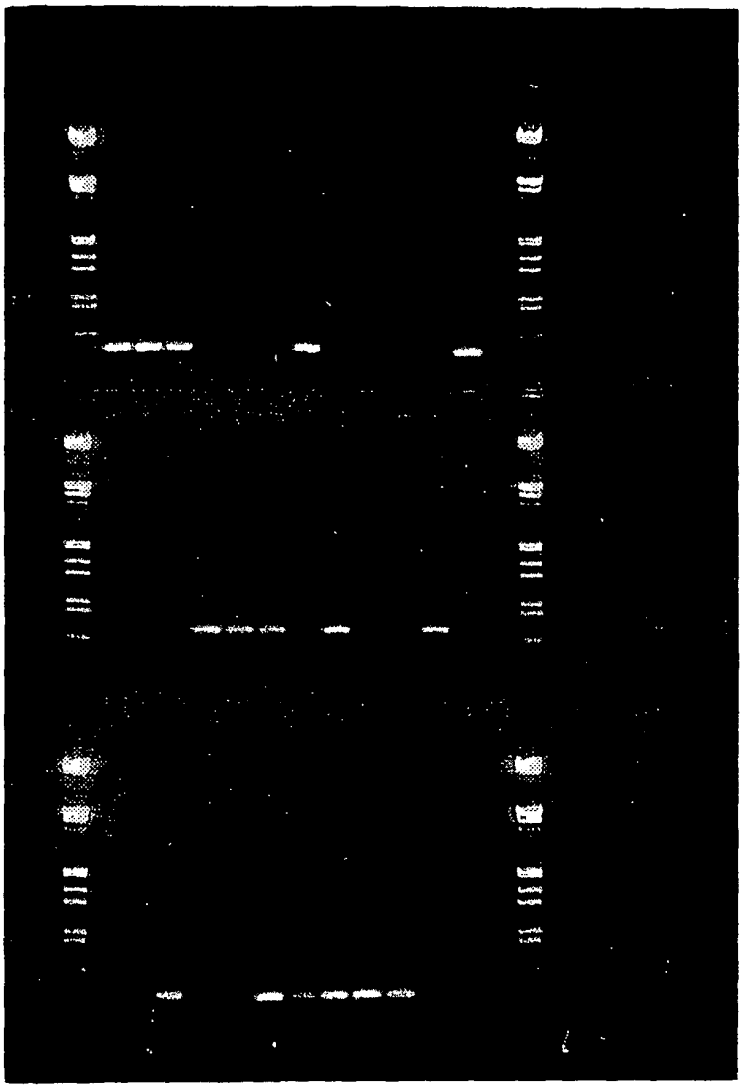

Fig. 2

AAGCTTGGAT ATTGATCACA TGATGGAGGT GATGGAAGCA TCTAAGTCTG CAGCGGGGTC
GGCGTGCCCA AGTCCGCAGG CTTATCAGGC AGCTTTTGAG GGAGCTGAGA
ACATTATCGT TGTGACGATT ACAGGTGGGC TATCGGGTAG TTTTAATGCG GCACGTGTAG
CTAGGGATAT GTATATCGAA GAGCATCCGA ATGTCAATAT CCATTTGATA
GATAGTTTGT CAGCCAGTGG GGAAATGGAT TTACTTGTAC ACCAAATCAA TCGCTTAATT
AGTGCAGGAT TAGATTTTCC ACAAGTAGTA GAAGCGATAA CTCACTATCG
GGAACACAGT AAGCTCCTCT TTGTTTTAGC GAAAGTTGAT AATCTTGTTA AGAATGGAAG
ACTGAGCAAA TTGGTAGGCA CTGTCGTTGG TCTTCTCAAT ATCCGTATGG
TTGGTGAGGC AAGTGCTGAA GGAAAATTAG AGTTGCTTCA AAAGGCGCGT GGTCATAAGA
AATCTGTGAC AGCAGCCTTT GAAGAAATGA AAAAAGCAGG CTATGATGGT
GGTCGAATTG TTATGGCCCA CCGCAACAAT GCTAAGTTCT TCCAACAATT CTCAGAGTTG
GTAAAAGCAA GTTTTCCAAC GGCTGTTATT GACGAAGTTG CAACATCAGG
TCTATGCAGT TTTTATGCTG AAGAAGGTGG ACTTTTGATG GGCTACGAAG TGAAAGCGTG
ATTCACAGAG TAATAATTTT GGGCTGTAAT TTCCGCTATA GAATAATCCC
CCTCTTCTTC TAAGTTCGAG GGGGATTGTT TGTATGAGAC TATTGGATTT CATTCATTCA
AATATCTTAC GAATTGCTCC AGTTTATCTG CAAAATCTTG TTCAAAGAAG
ATCTGTAAGA AATCAGCTTT CTGTCCGCTG AAATAATAAC ATTTTCCAAA CATGTGTTGG
ATGCTAGGAG AAAGAATCCC CTTGCTTAGC TGAAAGGTCA CGCTCCCCTT
TGGAATTCGA TACGGGATGT TTAAAGCGTA TTTCTCTAGA CAGTCTTTTA TTTTATTCCA
TTGAGCGTGA TAAATGTGAT GAAGATGCTG TGTGTTCCGC GCAAACATAC
CGTTATCAAT GTAGAGCGAG AGAGCTTTTT GCATGATAAG ATTGGTATCG TAGTCGATTA
GACTCTTATG TTTGATGAAG ATATCACGTA GCTGATTAGG AAGGCTGATT
GCACCGATTC GGAGGGCAGG AAAGAGTGTC GGTGTAAAAG ATTTTATATA GATGACGCGA
TTATCTGTAT CAAGATAGTG TAAAGGTAGG CTATGACTAG AGTCGAAATC
TGCTAAATAG TCATCCTCAA TGATGTAGAC ATCGTATTGC TTTGCTAATT TTACGATGGC
TGTTTTTGTT GCTATATCAT AGGTTGAACC GAGAGGGTTG TGCAAGCGAG
GAATTGTGTA GAAAAACTTA ATTTTTCCAG TTTGGAAGAT ACTTTCCAAT TCTTCTAGGT
CAATTCCATC TAAATTCCGT TCAATTGTTT GATAGGGGAT TCCTTGATGT
CGAATGAGCT CTATCATTCG TGAATAGGTA GGGTTCTCTA TCAAGATTTC CGTTTTTCCA
GCCAAGGTTT CCATTTGTGT GAGAATATAT AGAGCTTGTT GACTACCAGC
TGTGATAACC AGCTGGTCTT TTTTTGTATA GACATGATAG TCCATTAACA GACTTTGAAC
GGAGGAAATC AATTCTGCCA ATCCCTCTTG CTGGTGATAG TAGTTGAATA
GGTAATTTTC CCGCCCAATA AGACTTTCTT TTAGACAAAT CCGAAAATCT TCATAGGTAA
TTCTTGAAAG TCTGTAGGAT TGAGCTCTAC AGGTATGGTC TTGGAAATCT
CTATCCTCTA AGATATAATA ACCGCTTTTT TCGACAGCGT AGATCTTATT TTGGTATTTT
AATTCCAACA TAGCCTTTTG GACAGTGTCT TTGCTACAAT GATATTGCTC
GCGGAGTTGA CGGATAGAAG GTAATTTCTC TCCACGTTTG AATCGATGTT CCTCTATTCC
AGTCAAAATA TCTTGGATGA TAACTTGATA TTTTTTCATC TAGGTCCCCT
TTTTTATAGA CTATGTTACT AGCTAGTATA TAGAAAAAAT TGAAGAAAGA CAATATATGA
ATAATGGGGT TGAGGTTCAG GAATTAAGCT ACTCTATGGT ATAATTAAGT
GATGAAAATA ATTATACCTA ATGCAAAAGA AGTAAATACA AATCTAGAGA ATGCCTCGTT
TTATCTCCTG TCTGATCGAA GCAAGCCGGT GCTGGATGCC ATAAGTCAAT
TTGATGTAAA AAAGATGGCT GCCTTTTATA AATTGAATGA AGCAAAGGCT GAGTTAGAAG
CTGACCGTTG GTATCGAATC AGGACAGGTC AAGCAAAAAC CTATCCAGCC
TGGCAGTTAT ATGATGGTCT CATGTATCGT TATATGGATA GGCGAGGTAT AGATTCGAAA
GAAGAAAATT ATTTACGTGA CCACGTTCGT GTAGCGACAG CCTTATACGG
ATTGATTCAT CCTTTTGAAT TCATTTCACC TCACCGCTTA GATTTTCAAG GGAGCTTAAA
GATAGGCAAT CAGTCTTTGA AACAGTACTG GCGACCGTAT TATGACCAAG
AAGTTGGTGA TGATGAACTG ATTCTCTCAC TGGCTTCGTC AGAATTTGAG CAGGTGTTTT
CTCCCCAGAT TCAGAAAAGA TTAGTTAAAA TTCTTTTCAT GGAAGAAAAA
GCAGGTCAGC TAAAAGTTCA CTCGACTATA TCAAAAAAAG GCAGAGGAAG ATTGCTGTCC
TGGTTGGCTA AGAACAATAT TCAGGAATTA TCGGACATTC AAGATTTTAA
GGTGGATGGC TTTGAATATT GTACTTCCGA ATCAACGGCA AACCAACTTA CCTTCATACG
ATCAATAAAA ATGTGAAATT ATGAAAAAGA TAACGTTTTC CAGCGCTAAA
AAGGGTAGAA AAATATTAAT TTCTATGATA TAATGGATGC GTTATAGGTA AAAGTCTAGG
AAGGTTGTTT ATGAAAAAGA GAAGCGGACG AAGTAAGTCG TCCAAGTTCA
AATTGGTAAA TTTTGCGCTT TTGGGACTTT ATTCCATTAC TCTATGTTTG TTCTTAGTGA
CCATGTATCG CTATAACATC CTAGATTTCC GGTATTTAAA CTATATTGTG
ACGCTTTTGC TAGTAGGAGT GGCAGTATTG GCTGGATTAT TGATGTGGCG TAAGAAAGCG
CGCATATTTA CAGCGCTCTT ACTTGTTTTT TCACTGGTCA TCACGTCTGT

Fig. 3 DNA Serotype 2

TGGGATCTAT GGAATGCAAG AAGTTGTAAA ATTTTCAACA CGACTAAATT CAAATTCGAC
ATTTTCAGAA TATGAAATGA GTATCCTTGT CCCAGCAAAT AGTGATATTA
CGGACGTTCG TCAGCTTACT AGTATCCTTG CTCCAGCCGA ATACGACCAA GATAACATCA
CCGCTTTATT GGATGACATA TCCAAAATGG AATCTACTCA ACTAGCAACT
AGCCCCGGGA CTTCTTACCT GACAGCATAT CAATCTATGT TGAATGGCGA GAGTCAAGCG
ATGGTGTTCA ACGGAGTTTT TACCAATATT TTAGAAAATG AAGATCCAGG
CTTTTCTTCA AAAGTGAAAA AAATATATAG TTTCAAAGTG ACTCAGACTG TTGAAACAGC
TACTAAGCAG GTGAGTGGAG ATAGCTTTAA TATCTATATT AGTGGTATTG
ATGCTTATGG ACCGATTTCT ACGGTCTCTC GTTCAGATGT CAATATCATT ATGACTGTCA
ATCGTGCGAC ACATAAGATT TTATTGACAA CTACTCCACG AGATTCATAC
GTTGCTTTCG CAGATGGCGG GCAAAATCAA TACGATAAAC TAACACATGC TGGTATTTAC
GGTGTCAATG CTTCTGTGCA CACCTTAGAA AATTTTTATG GGATTGACAT
TAGCAATTAT GTGCGGTTGA ACTTCATTTC CTTCCTTCAA TTAATCGACT TGGTGGGTGG
AATTGATGTA TATAACGATC AAGAATTTAC AAGTTTACAT GGGAATTATC
ATTTCCCTGT TGGACAAGTT CATTTAAACT CAGACCAAGC ATTAGGCTTC GTTCGAGAGC
GCTACTCTTT AACAGGGGGT GACAATGACC GTGGTAAAAA CCAGGAAAAA
GTGATTGCTG CCTTGATTAA AAAGATGAGT ACGCCAGAGA ATCTAAAAAA TTACCAGGCA
ATCCTATCTG GATTGGAAGG CTCAATTCAA ACGGATTTGA GCTTAGAAAC
GATTATGAGT TTAGTGAATA CCCAACTAGA ATCAGGAACA CAATTTACAG TAGAGTCACA
AGCATTGACA GGAACAGGAC GCTCAGACTT ATCTTCTTAT GCGATGCCTG
GATCACAACT TTATATGATG GAAATTAACC AAGATAGTCT GGAGCAATCA AAGGCAGCGA
TTCAGTCCGT ACTTGTTGAA AAATAAAGAT TTTAGGAGAA AATATGAACA
ATCAAGAAGT AAATGCAATC GAAATCGATG TTTTATTCTT ACTAAAAACA ATTTGGAGAA
AGAAATTTTT AATTCTCTTA ACTGCAGTGT TGACTGCGGG GTTGGCATTT
GTCTACAGTA GTTTTTTAGT GACACCTCAA TATGACTCCA CTACCCGTAT CTATGTAGTG
AGTCAAAATG TTGAAGCCGG TGCGGGCTTG ACTAACCAAG AGTTACAAGC
GGGTACCTAT TTGGCAAAAG ACTATCGGGA AATTATCCTA TCACAAGATG TATTGACACA
AGTAGCAACG GAATTGAATC TGAAAGAGAG TTTGAAAGAA AAAATATCAG
TTTCTATTCC TGTTGATACT CGTATCGTTT CTATTTCTGT GCGTGATGCG GATCCAAATG
AAGCGGCACG TATTGCAAAT AGCCTTCGCA CCTTTGCAGT GCAAAAGGTT
GTTGAGGTCA CCAAGGTAAG CGATGTGACG ACACTTGAAG AAGCAGTCCC AGCGGAAGAA
CCAACCACTC CAAATACAAA ACGAAATATC TTGCTTGGTT TATTAGCTGG
AGGTATCTTG GCAACAGGTC TTGTACTGGT TATGGAGGTT TTGGATGACC GTGTAAAACG
TCCTCAGGAC ATCGAAGAGG TAATGGGATT GACATTGCTA GGTATAGTAC
CAGATTCGAA GAAATTAAAA TAGGAGAACA ATATGGCGAT GTTAGAAATT GCACGTACAA
AAAGAGAGGG AGTAAATAAA ACCGAGGAGT ATTTCAATGC TATCCGTACC
AATATTCAGC TTAGCGGAGC AGATATTAAG GTTGTTGGTA TTACCTCTGT TAAATCGAAT
GAAGGTAAGA GTACAACTGC GGCTAGTCTC GCTATTGCCT ATGCTCGTTC
AGGTTATAAG ACCGTCTTGG TGGATGCAGA TATCCGAAAT TCAGTCATGC CTGGTTTCTT
CAAGCCAATT ACAAAGATTA CAGGTTTGAC GGATTACCTA GCAGGGACAA
CAGACTTGTC TCAAGGATTA TGCGATACAG ATATTCCAAA CTTGACCGTA ATTGAGTCAG
GAAAGGTTTC TCCCAACCCT ACTGCCCTTT TACAAAGTAA GAATTTTGAA
AATCTACTTG CGACTCTTCG TCGCTATTAT GATTATGTTA TCGTTGACTG TCCACCATTA
GGACTGGTAA TTGATGCAGC TATCATTGCA CAAAAATGTG ATGCGATGGT
TGCAGTAGTA GAAGCAGGCA ATGTTAAGTG CTCATCTTTG AAAAAAGTAA AAGAGCAGTT
GGAACAAACA GGCACACCGT TCTTAGGCGT TATCTTGAAC AAATATGATA
TTGCCACTGA GAAGTATAGT GAATACGGAA ATTACGGCAA AAAAGCCTAA TTTCTCAGAT
AACATAAGTT TGATAAGTAG GTATTAATAT GATTGATATC CATTCGCATA
TCATATTTGG TGTGGATGAC GGTCCCAAAA CTATTGAAGA GAGCCTGAGT TTGATAAGCG
AAGCTTATCG TCAAGGTGTT CGCTATATCG TAGCGACATC TCATAGACGA
AAAGGGATGT TTGAAACACC AGAAAAAATC ATCATGATTA ACTTTCTTCA ACTTAAAGAG
GCAGTAGCAG AAGTTTATCC TGAAATACGA TTGTGCTATG GTGCTGAATT
GTATTATAGT AAAGATATCT TAAGCAAACT TGAAAAAAAG AAAGTACCAA CACTTAATGG
CTCGTGCTAT ATTCTCTTGG AGTTCAGTAC GGATACTCCT TGGAAAGAGA
TTCAAGAAGC AGTGAACGAA ATGACGCTAC TTGGGCTAAC TCCCGTACTT GCCCATATAG
AGCGTTATGA TGCTCTGGCA TTTCAGTCAG AGAGAGTAGA AAAGCTAATT
GACAAGGGAT GCTACACTCA GGTAAATAGT AACCATGTGT TGAAGCCTGC TTTAATTGGC
GAACGAGCAA AAGAATTTAA AAAACGTACT CGATATTTTT TAGAGCAGGA
TTTAGTACAT TGTGTTGCTA GCGATATGCA TAATTTATAT AGTAGACCTC CGTTTATGAG
GGAGGCGTAT CAGCTTGTAA AAAAAGAGTA TGGTGAGGAT AGAGCGAAGG

Fig. 3 cont.

CTTTGTTCAA GAAAAATCCT TTGTTGATAT TGAAAAATCA AGTACAGTAA CCTCATAGAA
ATAGTGGAGG AGCTATGAAT ATTGAAATAG GATATCGCCA AACGAAATTG
GCATTGTTTG ATATGATAGC AGTTACGATT TCTGCAATCT TAACAAGTCA TATACCAAAT
GCTGATTTAA ATCGTTCTGG AATTTTTATC ATAATGATGG TTCATTATTT
TGCATTTTTT ATATCTCGTA TGCCGGTTGA ATTTGAGTAT AGAGGTAATC TGATAGAGTT
TGAAAAAACA TTTAACTATA GTATAATATT TGTAATTTTT CTTATGGCAG
TTTCATTTAT GTTAGAGAAT AATTTCGCAC TTTCAAGACG TGGTGCCGTG TATTTCACAT
TAATAAACTT CGTTTTGGTA TACCTATTTA ACGTAATTAT TAAGCAGTTT
AAGGATAGCT TTCTATTTTC GACAACCTAT CAAAAAAAGA CGATTCTAAT TACAACGGCT
GAACTATGGG AAAATATGCA AGTTTTATTT GAATCAGATA TACTATTTCA
AAAAAATCTT GTTGCATTGG TAATTTTAGG TACAGAAATA GATAAAATTA ATTTACCATT
ACCGCTCTAT TATTCTGTTG AAGAAGCTAT AGGGTTTTCA ACAAGGGAAG
TGGTCGACTA CGTCTTTATA AATTTACCAA GTGAATATTT TGACTTAAAG CAATTAGTTT
CAGACTTTGA GTTGTTAGGT ATTGATGTAG GCGTTGATAT TAATTCATTC
GGTTTTACTG TGTTGAAGAA TAAAAAAATC CAAATGCTAG GTGACCATAG CATCGTCACT
TTTTCCACAA ATTTTTATAA GCCTAGTCAC ATCTGGATGA AACGACTTTT
AGATATACTT GGAGCAGTAG TCGGGTTAAT TATTAGTGGT ATAGTTTCTA TTTTGTTAAT
TCCAATTATT CGTAGAGATG GTGGGCCAGC CATTTTTGCT CAGAAACGAG
TTGGACAGAA TGGACGCATA TTTACATTCT ACAAGTTTCG TTCGATGTTT GTTGATGCCG
AGGTACGTAA GAAAGAATTA ATGGCTCAAA ACCAGATGCA AGGTGGGATG
TTCAAAATGG ACAACGATCC TAGAATTACT CCAATTGGAC ACTTCATACG AAAAACAAGT
TTAGATGAGT TACCACAATT TTATAATGTT CTAATTGGAG ATATGAGTCT
AGTCGGTACC CGTCCGCCTA CAGTTGATGA ATTTGAAAAA TATACTCCTA GTCAAAAGAG
AAGATTGAGT TTTAAACCAG GGATTACAGG TCTTTGGCAA GTGAGCGGAA
GAAGTGATAT CACAGATTTT AATGAAGTCG TTAGGCTGGA CCTAACATAC ATTGATAATT
GGACCATCTG GTCAGACATT AAGATTTTAT TGAAGACAGT GAAAGTTGTA
TTGTTGAGAG AGGGAGGTCA GTAAGACTCC TTTAAAACAA AGAATAGTAG TAGGGGATAT
GAGAACAGTT TATATTATTG GTTCAAAAGG AATACCAGCA AAGTATGGTG
GTTTCGAGAC TTTCGTAGAA AAATTAACTG AGTATCAGAA AGATAAATCA ATTAATTATT
TTGTTGCATG TACAAGAGAA AATTCAGCAA AATCAGATAT TACAGGAGAA
GTTTTTGAAC ATAATGGAGC AACATGTTTT AATATTGATG TGCCAAATAT TGGTTCAGCA
AAAGCCATTC TTTATGATAT TATGGCTCTC AAGAAATCTA TTGAAATTGC
CAAAGATAGA AATGATACCT CTCCAATTTT CTACATTCTT GCTTGTCGGA TTGGTCCTTT
CATTTATCTT TTTAAGAAGC AGATTGAATC AATTGGAGGT CAACTTTTCG
TAAACCCAGA CGGTCATGAA TGGCTACGTG AAAAGTGGAG TTATCCCGTC CGACAGTATT
GGAAATTTTC TGAGAGTTTG ATGTTAAAAT ACGCTGATTT ACTAATTTGT
GATAGCAAAA ATATTGAAAA ATATATTCAT GAAGATTATC GAAAATATGC TCCTGAAACA
TCTTATATTG CTTATGGAAC AGACTTAGAT AAATCACGCC TTTCTCCGAC
AGATAGTGTA GTACGTGAGT GGTATAAGGA GAAGGAAATT TCAGAAAATG ATTACTATTT
GGTTGTTGGA CGATTTGTGC CTGAAAATAA CTATGAAGTA ATGATTCGAG
AGTTTATGAA ATCATATTCA AGAAAAGATT TTGTTTTGAT AACGAATGTA GAGCATAATT
CCTTTTATGA GAAATTGAAA AAAGAAACAG GGTTCGATAA AGATAAGCGT
ATAAAGTTTG TTGGAACAGT CTATAATCAG GAGCTGTTAA AATATATTCG TGAAAATGCA
TTTGCTTATT TTCATGGTCA CGAGGTTGGA GGAACGAACC CATCTTTACT
TGAAGCACTT TCTTCTACTA AACTAAATCT TCTTCTAGAT GTGGGCTTTA ATAGAGAAGT
AGGGGAAGAA GGAGCGAAAT ACTGGAATAA AGATAATCTT CACAGAGTTA
TTGACAGTTG TGAGCAATTA TCACAAGAAC AAATTAATGA TATGGATAGT TTATCAACAA
AACAAGTCAA AGAAAGATTT TCTTGGGATT TTATTGTTGA TGAGTATGAG
AAGTTGTTTA AAGGATAAGT TATGAAAAAG ATTCTATATC TCCATGCTGG AGCAGAATTA
TATGGGGCAG ATAAGGTTCT CTTGGAACTT ATAAAAGGCT TAGATAAGAA
TGAATTTGAA GCGCATGTTA TCCTACCTAA TGATGGAGTC CTAGTGCCAG CATTAAGAGA
AGTTGGTGCG CAAGTTGAAG TTATTAACTA TCCAATTCTA CGTAGGAAAT
ATTTTAATCC AAAAGGGATT TTTGACTACT TCATATCATA TCATCACTAT TCTAAACAGA
TTGCTCAATA TGCCATAGAA AATAAGGTTG ACATAATTCA CAATAATACT
ACCGCTGTCT TAGAAGGCAT TTATCTGAAG CGAAAACTCA AATTACCTTT GTTGTGGCAT
GTTCATGAGA TTATTGTCAA ACCTAAATTC ATCTCTGATT CGATCAATTT
TTTAATGGGG CGTTTTGCTG ATAAGATTGT GACAGTTTCA CAGGCTGTGG CAAACCATAT
AAAACAATCA CCTCATATCA AAGATGACCA AATCAGTGTA ATCTACAATG
GGGTAGATAA TAAAGTGTTT TATCAGTCCG ATGCTCGGTC TGTTCGAGAA AGATTTGACA
TTGACGAAGA GGCTCTTGTC ATTGGTATGG TCGGTCGAGT CAATGCGTGG

Fig. 3 cont.

AAAGGACAAG GAGATTTTTT AGAAGCAGTT GCTCCTATAC TCGAACAGAA TCCAAAAGCT
ATCGCCTTTA TAGCAGGAAG TGCTTTTGAA GGAGAAGAGT GGCGAGTAGT
AGAATTAGAA AAGAAGATTT CTCAATTAAA GGTCTCTTCT CAAGTCAGAC GAATGGATTA
TTATGCAAAT ACCACTGAAT TATATAATAT GTTTGATATT TTTGTACTTC
CAAGTACTAA TCCAGACCCT CTACCAACGG TTGTACTAAA AGCAATGGCA TGCGGTAAAC
CTGTTGTCGG TTACCGACAT GGTGGTGTTT GTGAGATGGT GAAAGAAGGT
GTTAACGGTT TCTTAGTCAC TCCGAACTCA CCGTTAAATT TATCAAAAGT AATTCTTCAG
TTATCGGAAA ATATAAATCT CAGAAAAAAA ATTGGTAATA ATTCTATAGA
ACGTCAAAAA GAACATTTTT CGTTAAAAAG CTATGTAAAA AATTTTTCGA AAGTCTACAC
CTCCCTCAAA GTATACTGAT TGGCTGAAGT GAATGCTTTA GTATAGCGAT
TTATCGTATT CTCATTCGAT AAAACAAATG TTCAGAAACA GTTATAAGTT ATTTCTAAAG
GGCACCTCTA TAAACTCCCA AAATTGCGAA TTTGGAGTTA CGAAAGCCTT
GTTAAATCAA CATTTTAAAT TTTAGAAAAT TAGTTTTTAG AGCTCCCCTA AAATAGAAGA
TAACAGAAGG GAGCCTTCAA AAACTTCATT TTTAATTGGA TTGTAGAAAA
ACTGTTAAAT CAATATTTAG ATTTTTAGGA GTTCAGTTTT TGGGGGGAGA GCTTAATAAT
CTATGCACTA TATTTCGAAA AATATATGGT GTAAAATCAG AACTGATGGT
CGTGGCAAAA AAGAGAATGA GGAATTTATG AAAATTATTT CTTTTACAAT GGTTAATAAC
GAAAGTGAGA TAATAGAGTC ATTTATACGG TATAATTATA ACTTTATTGA
CGAGATGGTC ATTATTGATA ATGGTTGTAC AGATAACACG ATGCAAATTA TTTTTAATTT
GATTAAAGAG GGATATAAAA TATCCGTATA TGATGAGTCT TTAGAGGCAT
ATAATCAGTA TCGACTTGAT AATAAATATC TAACGAAAAT AATTGCTGAA AAAAATCCAG
ATTTGATAAT ACCTTTGGAT GCGGATGAAT TTTTAACAGC CGATTCAAAT
CCACGGAAAC TTTTGGAACA ACTGGACTTA GAAAAGATAC ATTATGTGAA TTGGCAATGG
TTTGTTATGA CTAAAAAAGA TGATATTAAT GATTCGTTTA TACCACGTAG
AATGCAATAT TGTTTTGAAA AACCTGTTTG GCATCATTCT GATGGTAAAC CAGTTACTAA
ATGTATAATT TCCGCTAAGT ATTACAAAAA AATGAATTTA AAGCTATCGA
TGGGACATCA CACTGTTTTT GGTAACCCAA ATGTAAGGAT AGAACATCAT AATGATTTGA
AATTTGCACA TTATCGAGCT ATTAGCCAAG AGCAATTAAT TTATAAAACA
ATTTGTTACA CTATTCGCGA TATTGCTACT ATGGAGAACA ATATCGAAAC AGCTCAAAGA
ACAAATCAGA TGGCGCTCAT TGAATCTGGC GTGGATATGT GGGAAACGGC
GAGAGAAGCC TCTTATTCAG GTTATGATTG TAATGTTATA CATGCACCAA TTGATTTAAG
TTTTTGTAAA GAAAATATTG TAATAAAATA TAACGAACTA TCCAGAGAAA
CAGTAGCAGA ACGCGTGATG AAAACGGGAA GAGAAATGGC TGTTCGTGCA TATAATGTGG
AGCGAAAACA AAAAGAAAAG AAATTTCTAA AACCTATTAT ATTTGTATTA
GATGGGTTAA AAGGAGATGA GTATATTCAT CCCAATCCAT CAAATCATTT GACGATCTTA
ACTGAAATGT ATAACGTCAG AGGCTTACTT ACCGATAATC ACCAAATTAA
ATTTCTCAAA GTTAATTATA GATTAATTAT AACTCCAGAT TTTGCTAAGT TTTTACCGCA
TGAATTTATT GTTGTACCAG ATACCTTGGA TATAGAGCAA GTTAAAAGCC
AGTATGTTGG TACAGGTGTA GACTTGTCAA AGATTATTTC TTTAAAAGAG TATCGAAAAG
AGATAGGCTT TATTGGTAAT TTGTATGCGC TTTTAGGATT TGTTCCGAAT
ATGCTCAATA GAATTTATCT ATATATTCAG AGAAACGGTA TTGCAAACAC TATTATAAAA
ATCAAGTCGA GATTGTGAGA GTTGTTTACT TTTATTTGTA ATTTTAAAAG
TAATGCAGGC AGATAGGAGA AAAACGTTTG GAAAAATGAG AATAAGAATT AATAATTTGT
TTTTTGTTGC CATAGCGTTT ATGGGCATAA TTATTAGTAA TTCGCAAGTT
GTTCTAGCGA TAGGCAAAGC TTCTGTGATT CAGTATCTAT CTTATTTAGT TTTGATTTTA
TGTATAGTTA ATGATTTATT AAAAAATAAC AAACATATTG TAGTTTATAA
ATTAGGGTAT TTGTTTCTTA TTATATTTTT ATTTACTATC GGAATATGTC AGCAAATTCT
TCCTATAACA ACTAAAATAT ATTTATCAAT TTCAATGATG ATTATTTCAG
TTTTAGCAAC GTTGCCAATA AGTTTGATAA AAGATATTGA TGATTTTAGA CGGATTTCAA
ATCATTTGTT ATTCGCTCTT TTTATAACTT CGATATTAGG AATAAAGATG
GGGGCAACGA TGTTCACGGG GGCAGTAGAA GGTATCGGTT TTAGTCAGGG TTTTAATGGA
GGATTGACGC ATAAGAACTT TTTTGGAATA ACTATTTTAA TGGGGTTCGT
ATTAACTTAC TTGGCGTATA AGTATGGTTC CTATAAAAGA ACGGATCGTT TTATTTTAGG
ATTAGAATTG TTTTTGATTC TTATTTCAAA CACACGCTCA GTTTATTTAA
TACTATTGCT TTTTCTATTT CTTGTTAATC TTGACAAAAT CAAAATAGAA CAAAGACAAT
GGAGTACGCT TAAATATATT TCCATGCTAT TTTGTGCTAT TTTTTTATAC
TATTTCTTTG GTTTTTTAAT AACACATAGT GATTCTTACG CTCATCGCGT TAATGGTCTT
ATTAATTTTT TTGAGTATTA TAGAAATGAT TGGTTCCATC TAATGTTTGG
TGCAGCGGAT TTGGCATATG GGGATTTAAC TTTAGACTAT GCTATAAGGG TTAGACGCGT
TTTAGGTTGG AATGGAACGC TTGAAATGCC CTTACTGAGT ATTATGTTAA

Fig. 3. cont.

AAAATGGTTT TATCGGTCTG GTAGGGTATG GGATTGTTTT ATATAAACTT TATCGTAATG
TAAGAATATT AAAAACAGAT AATATAAAAA CAATAGGAAA GTCTGTATTT
ATCATTGTAG TCCTATCTGC AACAGTAGAA AATTATATTG TAAATTTAAG TTTTGTATTT
ATGCCAATAT GTTTTTGTTT ATTAAATTCT ATATCTACTA TGGAATCAAC
TATTAACAAA CAACTGCAAA CATAAATTGG CAGGAATAGA GTTTTGAGTT GCTATTAATT
TGGTAGAGCA TATGTTCTAT AGGTGGCAAG ATAAAGATAG TATTTTTTAC
ATGATGATTT TTATGATAGC AAAGCAAGTT ACGGCATAAA AGGAATTAGA GGATGGAAAA
AGTCAGCATT ATTGTACCTA TTTTTAATAC GGAAAAGTAC TTAAGAGAGT
GTTTAGATAG CATTATTTCC CAATCGTATA CTAATCTAGA GATTCTTTTG ATAGATGACG
GTTCTTCAGA TTCATCAACG GATATATGTT TGGAATACGC AGAGCAAGAT
GGTAGAATAA AACTTTTCCG GTTACCAAAT GGTGGTGTTT CAAACGCAAG GAATTACGGT
ATCAAAAATA GCACAGCAAA TTATATTATG TTTGTAGATT CTGATGATAT
TGTTGACGGC AACATTGTTG AGTCCTTATA CACCTGTTTA AAAGAGAATG ATAGTGATTT
GTCGGGAGGG TTACTTGCTA CTTTTGATGG AAATTATCAA GAATCTGAGC
TGCAAAAGTG TCAAATTGAT TTGGAAGAGA TAAAAGAGGT GCGAGACTTA GGAAATGAAA
ATTTTCCCAA TCATTATATG AGCGGTATCT TTAATAGCCC TTGTTGCAAA
CTTTATAAGA ATATATATAT AAACCAAGGT TTTGACACTG AACAGTGGTT AGGAGAGGAC
TTATTATTTA ATCTAAATTA TTTAAAGAAT ATAAAAAAAG TCCGCTATGT
TAACAGAAAT CTTTATTTTG CCAGAAGAAG TTTACAAAGT ACTACAAATA CGTTTAAATA
TGATGTTTTT ATTCAATTAG AAAATTTAGA AGAAAAAACT TTTGATTTGT
TTGTTAAAAT ATTTGGTGGA CAATATGAAT TTTCTGTTTT TAAAGAGACG CTACAGTGGC
ATATTATTTA TTATAGCTTA TTAATGTTCA AAAATGGAGA TGAATCGCTT
CCAAAGAAAT TGCATATATT TAAGTATTTA TACAATAGGC ATTCTTTAGA TACTCTAAGT
ATTAAACGAA CGTCCTCTGT TTTTAAAAGA ATATGTAAAT TAATTGTTGC
TAATAATTTG TTTAAAATTT TTTTAAATAC TTTAATTAGG GAAGAAAAAA ATAATGATTA
ACATTTCTAT CATCGTCCCA ATTTACAATG TTGAACAATA TCTATCCAAG
TGTATAAATA GCATTGTAAA TCAGACCTAC AAACATATAG AGATTCTTCT GGTGAATGAC
GGTAGTACGG ATAATTCGGA AGAAATTTGT TTAGCATATG CGAAGAAAGA
TAGTCGCATT CGTTATTTTA AAAAAGAGAA CGGCGGGCTA TCAGATGCCC GTAATTATGG
CATAAGTCGC GCCAAGGGTG ACTACTTAGC TTTTATAGAC TCAGATGATT
TTATTCATTC GGAGTTCATC CAACGTTTAC ACGAAGCAAT TGAGAGAGAG AATGCCCTTG
TGGCAGTTGC TGGTTATGAT AGGGTAGATG CTTCGGGGCA TTTCTTAACA
GCAGAGCCGC TTCCTACAAA TCAGGCTGTT CTGAGCGGCA GGAATGTTTG TAAAAAGCTG
CTAGAGGCGG ATGGTCATCG CTTTGTGGTG GCCTGGAATA AACTCTATAA
AAAAGAACTA TTTGAAGATT TTCGATTTGA AAAGGGTAAG ATTCATGAAG ATGAATACTT
CACTTATCGC TTGCTCTATG AGTTAGAAAA AGTTGCAATA GTTAAGGAGT
GCTTGTACTA TTATGTTGAC CGAGAAAATA GTATCATAAC TTCTAGTATG ACTGACCATC
GCTTCCATTG CCTACTGGAA TTTCAAAATG AACGAATGGA CTTCTATGAA
AGTAGAGGAG ATAAAGAGCT CTTACTAGAG TGTTATCGTT CATTTTTAGC CTTTGCTGTT
TTGTTTTTAG GCAAATATAA TCATTGGTTG AGCAAACAGC AAAAGAAGCT
TCTCCAAACG CTATTTAGAA TTGTATATAA ACAATTGAAG CAAAATAAGC GACTTGCTTT
ACTAATGAAT GCTTATTATT TGGTAGGGTG TCTTCATCTT AATTTTAGTG
TCTTTCTGAA AACGGGGAAA GATAAAATTC AAGAAAGATT GAGAAGAAGT GAAAGTAGTA
CTCGGTAAGA ATGTTGTAAT AAATGGTTGA AAGAAAAGGG GATTAAAATG
AATCCAACAA ATAGTAGAAT AGCACTCTTT GATACGATTA AATGTATCAT GGTACTTTGT
GTTATTTTTA CACATCTGGA TTGGTCTGTT GAGCAGCGTC AATGGTTTAT
CTTTCCGTAT TTCGTTGACA TGGCTGTTCC AATTTTTCTG TTGCTTTCTG CCTATTTTCG
AACGAATAAG TGGAATACAA AACAAGAGAC GCTAAAGCTC AAGTTCAGCA
GTGGTATAAA AGAAAGTATA AACATGCTTT GTCTCTATGC TATCGTGATG GCTGTTAATG
TTTTATTGAG CTATTCGAGA ACCATCTGAT AGGAGTAAAG CCTTTTTCAG
GTTCTTCATC GCTCCGTTCA TTTGTCCTGT GGCTACTTTC TGGAGAATCG GGTCCAGGGA
GTTGGGAGTT ACTATGTTCC GTTGTTGATT CAGGTAGTTT TTTTATTACC
AATTTTGTAT GTTCTTTTCG AGAAAAATAA ATGGTTGGGC TTGCTTACTT GTTTTTTAGT
AAACTTTTCA GTGGATGCCA TATTTGCTAA CATGGCTGAA CACGGCATAT
ATATATAGAC TAATATCACT TCGTTATCTT TTTGTTCTAG GGCTTGGTTT TTTCTTTCAA
AGCAGGATGT GCGTTCCAAG GTAGATACTT TCATTGCGAC CCTATTTGGG
ATTATTGGAG CAATTCTGAT TTTTGTGAAT CATTCTATAG AGCCCTTCTC CTGGTTTTAT
GGTTGGAAGT CTACTTCCTT TCTATGCGTC CCATTTGCGT ATGCTATGCT
ATTTTTTATG ATAAAGTATG GACAGAAGAT TCCAGCAATA CTGTTGTCAA AATTGGGAGT
TGCTTCTTAT CATATCTACT TGACCCAGAT GCTGTATTTT TCAGTAGTCG

Fig. 3 cont.

CACCATTTTT AGCAGTGCAA TTTAAGGTAT CTTCGTTGAA TTTGTGGAAC GGCTTGTTTA
CCTTTCTAAT TTGCCTGTTT GGTGGCTATA TTTTCTACAA AGTGGATCTG
TTTATGAGAG TACGTGGAAA ACGATAATGA CTCATTTCAG ATTAGCAGAT GCCATTTCGT
TTATTAGCAG ATTCGCATGT TAATATTCCG ACAAAGAAAT TCAAATAGGT
TGACGAGAGA GGAGTGGTAT CTGTTTCTAA ACCCCAGTAT CCCCCTTTAT TTTCAAAGCT
ATATTTATTA ACTGAACAAG GAGAATTTTT AAGAGAACTG TTTGTTTAAT
CCCAGCACGA TCTGGTTCGA AAGGCTTACC GAATAAAAAC ATGCTATTTT TGGACGGGAA
ACCCATGATT TTTCACACGA TTGATGTGGC AATTGAATCA GGTTGTTTTG
AGAAAGAAGA CATCTATGTC AGTACGGATT CAGAAATGTA TAAGGGGGGC ACCTCTATAA
ATTCCCAAAA TTGCGAATTT GGAGTTACGA AAGCCTTGTT AAATCAACAT
CTTAAATTTT AGAAAATTAG TTTTTAGAGG TCCCCAAGGG GATTTGCGAG ACAAGAGGCA
TCAATGTATT GTTAAGACCC AAAGAACTAT CTACTTATCA TACTCCATCG
AATGAAGTCA GTACGCACTT TTTTACGAAT CTGGATTTTA TGAAGATTGT ATATTTGTTC
TTCTGCAAGT CACCTCACCG TTACGGACTG GCGAACAGAT AAAAGAAGCC
ATGAATATGT ACTTACAGGG GGACTCAGAA AATGTTTTGC ATTTCAATGA TGAAGGGCAA
GAAAGAGTGA ATCAGTACAT TATCGAAGCT GTACAGGGGT TATAAAAAGG
GGTTACTTAT CCTTAAAGTC TGTATGTAGA AGGAGAAAAA TTGAGACGAA TTTATATTTG
CCATACGATG TATCAGATCC TGATTTCCTT GTTAAAGATG GACGTTGAGA
GAGATAGTTT GATGTCCGTT GATATCATCG GGCATTTTCC AGATGTCAGG GAGCAACTGC
AGCAGCATGT TCATCTAATC GAGGGAGACG GAGCGTTCAT TTGATCTATA
TTCTTTGATA GCTAGATCAA AAACAAAAGA ACGCCTTTCC TTGTTACAGA GCTATGACGA
GGTGATCATT TTTCAAGATC ACCGTCAAGT CGGTCATTTT TTAAATAAAC
ATCGGATTCC CTATTCTCTT TTGGAGGATG GTTATAATTT TTTCAAGGAT AAAAGAGTGT
GCGATTTGGA GTCAATTCAA TCATCTGTCT GGAAAAGACT CTTTTATCAA
TGGTATTTTA AACCAACATA TTTGATTGGT TCAAGTCTCT ATTGTCAATC CATTGAGGTC
AATGATCTGT CGCTCGTACA ATTTGACTAG GCTTATAAAC CCTTTGTAGA
AGTTCCGAGA AAGCAATTAT TTGATCAAGC ATCGCCAGAG AAGGTGCAAG CGCTGCTGCA
GATATTTGGA GCAAGGGCGA TAGTAGCGGA TGAAGAGTCT TCTCAAAAAC
GATTGCTATT ATTGACCCAG CCCTTGTCTT GGGATTATCA TGTGACCGAA GAGAGTTGTT
GGAGATTTAT GTAGCAGGTC TTGCCCCTTA TCGGGAAGAC TATACAATCT
ACATAAAACC GCACCCACGA GATGGGGTTG ATTATTCATT TCTGGGTAAG GCTGTGGTGC
TTCTGCCTCA AGGTATTCCG TTTGAGTTGT TCGAAATGGC AGGTAATATC
CGTTTTGATA TCGGTATGAC CTATAGTTCG TCTGCTTTAG ATTTTTTAAA TTGTTTTGAA
GAGAAAGTGT ATTTAAAGGA CACTTTTCCT CTTCTTTCAA AAAATGATAT
TTTGCGTGAG GGGATAGAAT AGGAGGATTC ATGTCTAAAA AATCAATAGT TGTCTCAGGT
CTCGTCTATA CGATTGGAAC CATCCTCGTT CAGGGATTAG CCTTCATTAC
CCTCCCCATC TATACTCGTG TCATTTCTCA GGAAGTATAT GGGCAGTTTA GCTTGTATAA
TTCGTGGGTG GGGCTAGTTG GTCTCTTTAT CGGTCTACAG TTAGGTGGGG
CTTTTGGCCC GGGATGGGTA CACTTCCGCG AGAAATTTGA TGATTTCGTA TCCACCTTGA
TGGTCTCTTC TATCGCTTTC TTTTTACCAA TTTTTGGGCT ATCTTTTCTC
CTCAGTCAGC CCCTATCGCT CCTATTTGGT TTGCCTGATT GGGTCGTTCC GCTTTACTTT
TTGCAAAGTT TTATGAGTGT TGTGCAAGGA TTTTTTACGA CCTATTTAGT
GCAGCGGCAG CAGTCCATGT GGACTTTACT CCTATCGGTA CTGAGCGCTG TTATCAACAC
TGCTTTATCT TTATTTCTCA TCTTTTCGAT GGAGAATGAT TTCATCGCTC
GTGTAATGGC AAACTCGGCA ACGACTGGTG TTTTTGCTTG TGTGTCCTTG TTGTTTTTCT
ATAAGAAGAT TGGGCTTCAT TTTCGAAAGG ACTATCTTCG GTATGGTTTA
AGTATATCGA TTCCTCTTAT TTTTCATGGA TTAGGTCATA ATGTACTCAA TCAATTTGAC
AGAATCATGC TCGGCAAGAT GCTAACACTG TCAGATGTAG CCCTATACAG
TTTCGGCTAC ACACTTGCGT CTATCTTACA AATTGTGTTT TCGAGCTTGA ATACGGTATG
GTGTCCGTGG TATTTTGAGA AAAAGAGAGG TGCAGATAAA GATTTGCTCA
GTTATGTCCG TTACTATCTG GCGATTGGCC TGTTTGTGAC TTTTGGATTT CTAACAATTT
ACCCTGAATT AGCGATGTTG TTAGGTGGAT CTGAGTATCG TTTCAGTATG
GGATTTATTC CCATGATTAT TGTCGGGGTG TTCTTTGTAT TTCTTTATAG TTTTCCAGCC
AATATCCAGT TTTATAGTGG AAATACAAAG TTTTTGCCAA TTGGTACTTT
TATAGCAGGT GTACTAAATA TTTCCGTCCA CTTTGTTTTG ATACCGACAA AGAATTTATG
GTGCTGCTTT GCAACGACTG CTTCCTATCT GTTGTTGCTA GTCTTGCATT
ATTTTGTTGC TAAGAAAAAG TATGCTTACG ATGAAGTTGC GATTTCAACA TTTGTTAAGG
TAATTGCTCT TGTTGTCGTC TATACAGGCT TGATGACAGT ATTTGTCGGT
TCAATCTGGA TTCGTTGGTC ACTAGGAATA GCGGTTCTAG TCGTTTATGC CTACATTTTT
AGAAAGGAAT TAACAGTTGC CCTCAATACA TTCAGGGAAA AACGGTCTAA

Fig. 3 cont.

ATAAGGGCAC CTCTATAAAC TCCCAAAATT GCGAATTTGG AGTTACGAAA GCCTTGTTAA
ATCAAACATT TTAAATTTTA GAAAATTAGT TTTTAGAGGT CCCCATATAA
AAACGTCCCA AATGAGAGGT GCTCATAAGA ATTGACCATC ACTGCCATCT ACCCAAAGTT
CAAGTATTCT CTACCATGAA AATTGTGCTA TAATCAAGTA TAAAGAAGGG
AATGTTTCTT AAAGGACGTA TGCGCCTCTG CTTATGCCAG AAGTCATGAG GTAAATCTCC
CTAAAAATTG GGTAGAAAAG CAGATTAAAC TTCCACCAAT CTATTGAAGA
TCGTGTTGAA GAGCAGGCTT TAGAAGCAAC AAGCCCTGAG ACTATTCGAA AGAAATCTAG
GGCTATTTTT TCTAATCGGC TATCAGAAGT GAAGTAGCGA TCTTTATTAG
TGTTCTTTTA CTACTTAAGG AAAACCAAGC TGCTCCCTCA AGACTTTATG GGAGCGATTT
ACAGTCATTT TTAGAAAGGA AATAAAATGG TTTATATTAT TGCAGAAATT
GGTTGTAATC ACAACGGTGA TGTTCATCTA GCACGGAAAA TGGTAGAAGT TGCCGTTGAT
TGTGGTGTGG ATGCCGTTAA ATTTCAGACA TTTAAGGCAG ATTTGTTGAT
TTCAAAATAC GCACCAAAGG CCGAATACCA AAAAATTACA ACAGGAGAGT CAGATTCTGA
GCTCGAAATG ACTCGTCGTT TGGAATTGAG CTTTGAAGAG TATCTTGATT
TGCGTGATTA CTGTCTTGAA AAGGGAGTTG ATGTGTTTTC GACACCTTTT GATGAGGAAT
CATTGGACTT CTTGATTAGC ACAGATATGC CCGTTTATAA GATTCCATCT
GGTGAGATTA CCAATCTTCC CTATTTGGAA AAAATTGGTC GTCAAGCTAA GAAAGTTATT
CTTTCAACTG GTATGGCTGT TATGGATGAA ATTCATCAAG CGGTGAAGAT
TTTGCAGGAA AATGGAACGA CCGATATTTC GATTTTGCAT TGTACAACCG AGTATCCAAC
CCCTTACCCT GCTTTGAATT TGAATGTCTT GCATACCTTG AAAAAAGAAT
TTCCAAACTT AACAATTGGC TATTCAGACC ATAGTGTTGG TTCAGAAGTA CCCATCGCTG
CTGCAGCAAT GGGAGCTGAA TTGATTGAAA AGCACTTTAC TCTGGACAAT
GAAATGGAAG GACCAGATCA TAAAGCGAGT GCTACTCCTG ATATCTTAGC AGCCTTGGTA
AAAGGAGTGA GGATAGTGGA ACAATCTCTT GGTAAATTTG AAAAAGAGCC
AGAAGAAGTT GAAGTACGAA ATAAAATTGT AGCTAGAAAA TCTATTGTTG CCAAAAAAGC
AATTGCTAAA GGCGAAGTCT TTACAGAAGA AAACATCACT GTCAAAAGAC
CAGGAAATGG AATTTCGCCA ATGGAATGGT ACAAAGTCTT GGGGCAGGTG AGTGAGCAGG
ATTTTGAGGA AGACCAAAAT ATTTGCCATA GTGCTTTTGA AAATCAAATG
TAAGCGGAGT AAGGATGAAA AAAATTTGTT TTGTGACAGG CTCTCGTGCC GAATATGGGA
TTATGCGTCG CTTATTGAGC TATCTACAGG ATGATCCAGA AATGGAGCTG
GATCTTGTAG TGACAGCCAT GCATCTAGAA GAAAAATATG GGATGACGGT CAAAGACATC
GAAGCGGACA AGCGTAGGAT TGTCAAGCGG ATTCCATTGC ATTTGACGGA
TACGTCTAAG CAGACAATCG TCAAATCTTT AGCGACCTTG ACAGAGCAAC TCACGGTTCT
TTTTGAAGAA GTCCAGTATG ACTTGGTGTT GATTCTGGGG GATCGCTATG
AGATGCTACC AGTTGCCAAT GCTGCGTTGC TTTATAATAT TCCTATTTGC CATATTCATG
GTGGTGAAAA AACCATGGGA AATTTTGATG AGTCGATTCG CCATGCCATT
ACCAAGATGA GTCACCTTCA TCTGACATCA ACGGATGAAT TTAGAAATCG TGTCATTCAA
CTAGGAGAAA ATCCAACCAT GTACTGAACA TCGGAGCTAT GGGTGTTGAA
AATGTTTTAA AACAAGACTT TTTGACAAGA GAAGAGTTGG CGATGGAACT TGGAATTGAT
TTTGCCGAGG ATTACTATGT TGTACTCTTT CACCCTGTTA CCTTGGAGGA
TAACACAGCC GAAGAACAAA CGCAGGCCTT ATTAGATGCT CTAAAAGAAG ATGGTAGCCA
GTGTTTGATA ATTGGATCCA ATTCGGATAC ACATGCCGAT AAGATAATGG
AATTGATGCA TGAATTTGTA AAACAAGACT CTGATTCTTA CATCTTTACT TCGCTTCCAA
CTCGTTATTA CCATTCCTTG GTCAAGCATT CACAAGGTTT AATAGGGAAT
TCTTCGTCAG GTTTGATTGA AGTGCCCTCA TTACAGGTTC CGACCTTAAA TATTGGAAAT
CGCCAATTTG GACGTTTGTC AGGACCGAGT GTGGTACATG TTGGAACTTC
TAAGGAAGCG ATTGTTGGTG GTTTGGGGCA ATTACGTGAT GTGATAGATT TTACCAATCC
ATTTGAACAA CCTGATTCTG CTTTACAAGG TTATCGAGCT ATCAAGGAAT
TTTTATCTGT ACAGGCCTCA ACCATGAAAG AGTTTTATGA TAGATAGGGG AGAAAGTTTG
ATGAAAAAAG TAGCCTTTCT AGGAGCGGGT ACCTTTTCAG ATGGTGTCCT
TCCTTGGTTG GATAGAACTC GATATGAACT CATTGGATAT TTTGAAGATA AACCGATCAG
TGACTATCGT GGCTATCCTG TATTTGGTCC CTTGCAAGAT GTCCTAACCT
ATTTGGATGA TGGAAAAGTA GATGCTGTCT TCGTCACTAT AGGTGACAAT GTCAAGCGCA
AGGAAATCTT TGACTTGCTT GCCAAAGATC ATTATGATGC TTTGTTCAAC
ATCATTAGCG AGCAAGCCAA TATTTTTTCC CCAGATAGTA TCAAGGGACG AGGGGTTTTC
ATAGGTTTTT CAAGTTTTGT AGGAGCCGAT TCCTATGTCT ATGACAATTG
TATCATCAAT ACGGGTGCCA TTGTGGAACA TCATACCACG GTGGAGGCCC ATTGTAACAT
TACTCCAGGA GTGACCATAA ATGGCTTGTG CCGTATCGGA GAAAGCACTT
ATATTGGAAG TGGTTCAACA GTGATTCAAT GTATCGAGAT TGCACCTTAT ACAACATTGG
GGGCAGGGAC AGTTGTTTTG AAATCGTTGA CGGAGTCAGG GACCTATGTT

Fig. 3 cont.

GGTGTACCTG CTAGAAAGAT TAAATAGGTG AATTGATGGA ACCAATTTGT CTGATTCCTG
CTCGGTCAGG ATCAAAAGGT TTACCAAATA AAAACATGTT ATTTTTAGAT
GGTGTACCGA TGATTTTCCA TACCATTCGA GCTGCGATTG AGTCTGGATG TTTTAAGAAA
GAAAATATAT ATGTCAGTAC TGATTCAGAG GTTTACAAGG AAATTTGTGA
AACAACTGGG GTTCAAGTCC TCATGCGTCC AGCTGACTTG GCGACAGATT TTACAACCTC
TTTTCAACTG AACGAACATT TTTTACAAGA TTTTTCTGAT GACCAAGTAT
TTGTTCTCCT GCAAGTTACG TCCCCATTAA GATCGGGAAA ACATGTCAAG GAGGCGATGG
AGTTATATGG GAAAGGTCAA GCTGACCACG TTGTTAGCTT TACCAAAGTC
GATAAGTCTC CAACATTGTT TTCAACTTTA GACGAAAACG GATTCGCTAA GGATATTGCA
GGATTAGGTG GCAGTTATCG TCGTCAAGAT GAGAAAACAC TCTACTATCC
TAATGGAGCG ATTTATATTT CTTCTAAGCA GGCTTATTTA GCGGATAAAA CTTATTTTTC
TGAAAAAACA GCGGCCTATG TGATGACGAA GGAAGATTCG ATTGATGTAG
ATGATCACTT TGATTTTACT GGTGTTATTG GTCGAATTTA CTTTGATTAC CAGCGTCGTG
AGCAACAAAA CAAACCATTT TATAAAAGAG AGTTAAAGCG TTTATGTGAG
CAACGAGTCC ATGATAGTCT TGTGATTGGC GATAGTCGTC TGTTAGCCTT GTTACTGGAT
GGTTTCGATA ATATCAGCAT CGGTGGGATG ACAGCTTCGA CAGCACTTGA
AAACCAAGGT CTCTTTTTGG CTACTCCGAT AAAGAAAGTT TTGCTTTCTC TTGGTGTGAA
TGATTTGATT ACTGACTATC CCTTGCATAT GATTGAGGAT ACTATTCGCC
AGCTGATGGA AAGTCTTGTT TCCAAAGCAG AGCAGGTTTT TGTGACGACG ATTGCCTACA
CGCTGTTTCG TGATAGCGTT TCCAATGAAG AAATTGTGCA GCTGAATGAC
GTTATTGTTC AGTCAGCAAG TGAACTGGGT ATTTCAGTGA TTGATCTAAA TGAAGTTGTT
GAAAAAGAGG CGATGCTTGA CTATCAGTAT ACCAATGATG GATTGCATTT
CAATCAGATT GGACAAGAGC GTGTGAATCA GCTGATTTTG ACAAGTTTGA CAAGATAATT
TGGTGATAGA AGCTATTTCA GTGGCTAGAC TATGTTGGTA TGTGTTTTAG
AGCCCAGGAA TAACATCTGT AGAGGATGCT AGCCTTGAGA ATTGACAACC ATTTAGTTGT
TTTAATTATA TAAGGGGACC TCTAAAAACT CCCTAAATTT CCCAAAAATG
AGATAATAGA ATAAAAAGTA ATGAGGAGAG CTGTCATGCA TTTATTCACA GACGATGAAA
AAATCTTGTC AAAACTATCA GAGAAAGGCA ATCCCTTAGA ACGTTTGGAT
GCCGTTATGG ATTGGAATAT CTTTCTTCCA TTGTTGTCAG AGTTATTCAG TCGTAAAGAT
AAAGTCATCA GTCGTGGCGG TCGTCCTCAC CTAGACTATC TCATGATGTT
CAAAGCGCTC TTGCTTCAAC GTCTTCATAA CCTATCTGAC GATGCCATGG AATATCAACT
GCTGGATCGT ATATCTTTTC GTCGTTTTGT TGGTTGTCAT GAAGACACTG
TTCCCGATGC GAAAACTATC TGGCTCTATC GTGAGAAATT AACCAAGTCA GGTCGTGAAA
AGGAGTTGTT CGATTTGTTC TATGCCCATC TCACAGATGA AGGGGTGATT
GCCCATTCAG GTCAGATTGT GGATGCTACC TTTGTCGAAT GCCCTAAACA ACGCAATTCA
CGTGAGGACA ATCAGAAAAT CAAAACTTAT CGAAAATTAT GAGGTCACAA
CAGCTAGTGT ACACGACTCC AATGTCCTAG CTCCTCTTTG TGATGCCAAT GAAGCGGTTT
TTGATGACAG TGCTTATGTT GGAAAATCAG TACCAGAAGG TTGTCGCCAC
CACACGATTC GTCGTGCTTT TAGAAATAAA CCGTTGACTG AGACTGATAA GGTCATTAAT
CGACATATTA CCAAAGTCCG TTGTCGCGTT GAGCATGGTT TTGGCTTCAT
TGAAACTAAC ATGAAAGGTA ACATCTGTCG AGCAATTGGG AAGGCACGAG CTGAAACCAA
TGTGACCTTA ACCAACCTGC TCTACAATAT CTGTCGTTTT GAGCAAATCA
AACGACTGGG ATTACCATCC GTGGGCTTAG TGCGCCCAAA AAATAGGAAA ATAAGCAAAA
AGAGGCTGGG CAAAAACTAG TTTCTCACAA TAAAAAAACG GCTCTTTGTC
AACTGTAGTG GGTAGACGAA AAGCTAACAC CTAGAGAGGA CGAAATTCGT TCTCTCATTT
TTGATGTTTA AAGCGTAACC GCCTAATAAC AAGGTATCTA TCCAATCACA
CATTCCTCCA TTATATAGTT AAATGAAACA AAAACAGTAC ATCTATGATA TAATGTATTT
ATGGCATATT CATTAGATTT TCGTAAAAAA GTTCTCGCAT ACTGTGAGAA
AACCGGCAGT ATTACTGAAG CATCAGCTAT TTTCCAAGTT TCACGTAACA CTATCTATCA
ATGGCTAAAA TTAAAAGAGA AAACCGGCGA GCTTCATCAC CAAGTTAAAG
GAACCAAGCC AAGAAAAGTG GATAGAGATA AATTAAAGAA TTATCTTGAA ACTCATCCAG
ATGCTTATTT GACTGAAATA GCTTCTGAAT TTGACTGTCA TCCAACAGCT
ATTCATTACC CCCTCAAAGC TATGGGATAT ACTCGAAAAA AAAGAGCTGT ACCTACTATG
AACAAGACCC TGAAAAAGTA GAACTGTTCC TTAAAGAATT GAATAACTTA
AGCCACTTGA CTCCTGTTTA TATTGACGAG ACAGGGTTTG AGACATATTT TCATCGAAAA
TATGGTCGCT CTTTGAAAGG TCAGTTGATA AAAGGTAAGG TCTCTGGAAG
AAGATACCAG CGGATATCTT TAGTAGCAGG TCTCATAAAT GGTGCGCTTA TAGCCCCGAT
GACATACAAA GATACTATGA CGAGTGGCTT TTTCGAAGCT T

Fig. 3 cont.

SLDIDHMMEVMEASKSAAGSACPSPQAYQAAFEGAENIIVVTITGGLSGSFNAARVARDM
YIEERPNVNIHLIDSLSASGEMDLLVHQINRLISAGLDFPQVVEAITHYREHSKLLFVLA
KVDNLVKNGRLSKLVGTVVGLLNIRMVGEASAEGKLELLQKARGHKKSVTAAFEEMKKAG
YDGGRIVMAHRNNAKFFQQFSELVKASFPTAVIDEVATSGLCSFYAEEGGLLMGYEVKA

Fig. 3 cont. ORF2Z

MKKYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQKAMLELKYQNKIYAVE
KSGYYILEDRDFQDHTCRAQSYRLSRITYEDFRICLKESLIGRENYLFNYYHQQEGLAEL
ISSVQSLLMDYHVYTKKDQLVITAGSQQALYILTQMETLAGKTEILIENPTYSRMIELIR
HQGIPYQTIERNLDGIDLEELESIFQTGKIKFFYTIPRLHNPLGSTYDIATKTAIVKLAK
QYDVYIIEDDYLADFDSSHSLPLHYLDTDNRVIYIKSFTPTLFPALRIGAISLPNQLRDI
FIKHKSLIDYDTNLIMQKALSLYIDNGMFARNTQHLHHIYHAQWNKIKDCLEKYALNIPY
RIPKGSVTFQLSKGILSPSIQHMFGKCYYFSGQKADFLQIFFEQDFADKLEQFVRYLNE

Fig. 3 cont. ORF2Y

MKIIIPNAKEVNTNLENASFYLLSDRSKPVLDAISQFDVKKMAAFYKLNEAKAELEADRW
YRIRTGQAKTYPAWQLYDGLMYRYMDRRGIDSKEENYLRDHVRVATALYGLIHPFEFISP
HRLDFQGSLKIGNQSLKQYWRPYYDQEVGDDELILSLASSEFEQVFSPQIQKRLVKILFM
EEKAGQLKVHSTISKKGRGRLLSWLAKNNIQELSDIQDFKVDGFEYCTSESTANQLTFXR
SIKM

Fig. 3 cont. ORF2X

MKKRSGRSKSSKFKLVNFALLGLYSITLCLFLVTMYRYNILDFRYLNYIVTLLLVGVAVL
AGLLMWRKKARIFTALLLVFSLVITSVGIYGMQEVVKFSTRLNSNSTFSEYEMSILVPAN
SDITDVRQLTSILAPAEYDQDNITALLDDISKMESTQLATSPGTSYLTAYQSMLNGESQA
MVFNGVFTNILENEDPGFSSKVKKIYSFKVTQTVETATKQVSGDSFNIYISGIDAYGPIS
TVSRSDVNIIMTVNRATHKILLTTTPRDSYVAFADGGQNQYDKLTHAGIYGVNASVHTLE
NFYGIDISNYVRLNFISFLQLIDLVGGIDVYNDQEFTSLHGNYHFPVGQVHLNSDQALGF
VRERYSLTGGDNDRGKNQEKVIAALIKKMSTPENLKNYQAILSGLEGSIQTDLSLETIMS
LVNTQLESGTQFTVESQALTGTGRSDLSSYAMPGSQLYMMEINQDSLEQSKAAIQSVLVE
K

Fig. 3 cont. CPS2A

MNNQEVNAIEIDVLFLLKTIWRKKFLILLTAVLTAGLAFVYSSFLVTPQYDSTTRIYVVS
QNVEAGAGLTNQELQAGTYLAKDYREIILSQDVLTQVATELNLKESLKEKISVSIPVDTR
IVSISVRDADPNEAARIANSLRTFAVQKVVEVTKVSDVTTLEEAVPAEEPTTPNTKRNIL
LGLLAGGILATGLVLVMEVLDDRVKRPQDIEEVMGLTLLGIVPDSKKLK

Fig. 3 cont. CPS2B

MAMLEIARTKREGVNKTEEYFNAIRTNIQLSGADIKVVGITSVKSNEGKSTTAASLAIAY
ARSGYKTVLVDADIRNSVMPGFFKPITKITGLTDYLAGTTDLSQGLCDTDIPNLTVIESG
KVSPNPTALLQSKNFENLLATLRRYYDYVIVDCPPLGLVIDAAIIAQKCDAMVAVVEAGN
VKCSSLKKVKEQLEQTGTPFLGVILNKYDIATEKYSEYGNYGKKA

Fig. 3 cont. CPS2C

MIDIHSHIIFGVDDGPKTIEESLSLISEAYRQGVRYIVATSHRRKGMFETPEKIIMINFL
QLKEAVAEVYPEIRLCYGAELYYSKDILSKLEKKKVPTLNGSCYILLEFSTDTPWKEIQE
AVNEMTLLGLTPVLAHIERYDALAFQSERVEKLIDKGCYTQVNSNHVLKPALIGERAKEF
KKRTRYFLEQDLVHCVASDMHNLYSRPPFMREAYQLVKKEYGEDRAKALFKKNPLLILKN
QVQ

Fig. 3 cont. CPS2D

MNIEIGYRQTKLALFDMIAVTISAILTSHIPNADLNRSGIFIIMMVHYFAFFISRMPVEF
EYRGNLIEFEKTFNYSIIFVIFLMAVSFMLENNFALSRRGAVYFTLINFVLVYLFNVIIK
QFKDSFLFSTTYQKKTILITTAELWENMQVLFESDILFQKNLVALVILGTEIDKINLPLP
LYYSVEEAIGFSTREVVDYVFINLPSEYFDLKQLVSDFELLGIDVGVDINSFGFTVLKNK
KIQMLGDHSIVTFSTNFYKPSHIWMKRLLDILGAVVGLIISGIVSILLIPIIRRDGGPAI
FAQKRVGQNGRIFTFYKFRSMFVDAEVRKKELMAQNQMQGGMFKMDNDPRITPIGHFIRK
TSLDELPQFYNVLIGDMSLVGTRPPTVDEFEKYTPSQKRRLSFKPGITGLWQVSGRSDIT
DFNEVVRLDLTYIDNWTIWSDIKILLKTVKVVLLREGGQ

Fig. 3 cont. CPS2E

MRTVYIIGSKGIPAKYGGFETFVEKLTEYQKDKSINYFVACTRENSAKSDITGEVFEHNG
ATCFNIDVPNIGSAKAILYDIMALKKSIEIAKDRNDTSPIFYILACRIGPFIYLFKKQIE
SIGGQLFVNPDGHEWLREKWSYPVRQYWKFSESLMLKYADLLICDSKNIEKYIHEDYRKY
APETSYIAYGTDLDKSRLSPTDSVVREWYKEKEISENDYYLVVGRFVPENNYEVMIREFM
KSYSRKDFVLITNVEHNSFYEKLKKETGFDKDKRIKFVGTVYNQELLKYIRENAFAYFHG
HEVGGTNPSLLEALSSTKLNLLLDVGFNREVGEEGAKYWNKDNLHRVIDSCEQLSQEQIN
DMDSLSTKQVKERFSWDFIVDEYEKLFKG

Fig. 3 cont. CPS2F

MKKILYLHAGAELYGADKVLLELIKGLDKNEFEAHVILPNDGVLVPALREVGAQVEVINY
PILRRKYFNPKGIFDYFISYHHYSKQIAQYAIENKVDIIHNNTTAVLEGIYLKRKLKLPL
LWHVHEIIVKPKFISDSINFLMGRFADKIVTVSQAVANHIKQSPHIKDDQISVIYNGVDN
KVFYQSDARSVRERFDIDEEALVIGMVGRVNAWKGQGDFLEAVAPILEQNPKAIAFIAGS
AFEGEEWRVVELEKKISQLKVSSQVXRMDYYANTTELYNMFDIFVLPSTNPDPLPTVVLK
AMACGKPVVGYRHGGVCEMVKEGVNGFLVTPNSPLNLSKVILQLSENINLRKKIGNNSIE
RQKEHFSLKSYVKNFSKVYTSLKVY

Fig. 3 cont. CPS2G

MKIISFTMVNNESEIIESFIRYNYNFIDEMVIIDNGCTDNTMQIIFNLIKEGYKISVYDE
SLEAYNQYRLDNKYLTKIIAEKNPDLIIPLDADEFLTADSNPRKLLEQLDLEKIHYVNWQ
WFVMTKKDDINDSFIPRRMQYCFEKPVWHHSDGKPVTKCIISAKYYKKMNLKLSMGHHTV
FGNPNVRIEHHNDLKFAHYRAISQEQLIYKTICYTIRDIATMENNIETAQRTNQMALIES
GVDMWETAREASYSGYDCNVIHAPIDLSFCKENIVIKYNELSRETVAERVMKTGREMAVR
AYNVERKQKEKKFLKPIIFVLDGLKGDEYIHPNPSNHLTILTEMYNVRGLLTDNHQIKFL
KVNYRLIITPDFAKFLPHEFIVVPDTXDIEQVKSQYVGTGVDLSKIISLKEYRKEIGFIG
NLYALLGFVPNMLNRIYLYIQRNGIANTIIKIKSRL.

Fig. 3 cont. CPS2H

MQADRRKTFGKMRIRINNLFFVAIAFMGIIISNSQVVLAIGKASVIQYLSYLVLILCIVN
DLLKNNKHIVVYKLGYLFLIIFLFTIGICQQILPITTKIYLSISMMIISVLATLPISLIK
DIDDFRRISNHLLFALFITSILGIKMGATMFTGAVEGIGFSQGFNGGLTHKNFFGITILM
GFVLTYLAYKYGSYKRTDRFILGLELFLILISNTRSVYLILLLFLFLVNLDKIKIEQRQW
STLKYISMLFCAIFLYYFFGFLITHSDSYAHRVNGLINFFEYYRNDWFHLMFGAADLAYG
DLTLDYAIRVRRVLGWNGTLEMPLLSIMLKNGFIGLVGYGIVLYKLYRNVRILKTDNIKT
IGKSVFIIVVLSATVENYIVNLSFVFMPICFCLLNSISTMESTINKQLQT

Fig. 3 cont. CPS2I

MEKVSIIVPIFNTEKYLRECLDSIISQSYTNLEILLIDDGSSDSSTDICLEYAEQDGRIK
LFRLPNGGVSNARNYGIKNSTANYIMFVDSDDIVDGNIVESLYTCLKENDSDLSGGLLAT
FDGNYQESELQKCQIDLEEIKEVRDLGNENFPNHYMSGIFNSPCCKLYKNIYINQGFDTE
QWLGEDLLFNLNYLKNIKKVRYVNRNLYFARRSLQSTTNTFKYDVFIQLENLEEKTFDLF
VKIFGGQYEFSVFKETLQWHIIYYSLLMFKNGDESLPKKLHIFKYLYNRHSLDTLSIKRT
SSVFKRICKLIVANNLFKIFLNTLIREEKNND

Fig. 3 cont. CPS2J

MINISIIVPI YNVEQYLSKC INSIVNQTYK HIEILLVNDG STDNSEEICL AYAKKDSRIR
YFKKENGGLS DARNYGISRA KGDYLAFIDS DDFIHSEFIQ RLHEAIEREN
ALVAVAGYDR VDASGHFLTA EPLPTNQAVL SGRNVCKKLL EADGHRFVVA WNKLYKKELF
EDFRFEKGKI HEDEYFTYRL LYELEKVAIV KECLYYYVDR ENSIITSSMT
DHRFHCLLEF QNERMDFYES RGDKELLLEC YRSFLAFAVL FLGKYNHWLS KQQKKLLQTL
FRIVYKQLKQ NKRLALLMNA YYLVGCLHLN FSVFLKTGKD KIQERLRRSE
SSTR

Fig. 3 cont. CPS2K

MSKKSIVVSG LVYTIGTILV QGLAFITLPI YTRVISQEVY GQFSLYNSWV GLVGLFIGLQ
LGGAFGPGWV HFREKFDDFV STLMVSSIAF FLPIFGLSFL LSQPLSLLFG
LPDWVVPLIF LQSLMIVVQG FFTTYLVQRQ QSMWTLPLSV LSAVINTALS LFLTFPMEND
FIARVMANPA TTGVLACVSX WFSQKKNGLH FRKDYLRYGL SISIPLIFHG
LGHNVLNQFD RIMLGKMLTL SDVALYSFGY TLASILQIVF SSLNTVWCPW YFEKKRGADK
DLLSYVRYYL AIGLFVTFGF LTIYPELAML LGGSEYRFSM GFIPMIIVGV
FFVFLYSFPA NIQFYSGNTK FLPIGTFIAG VLNISVHFVL IPTKNLWCCF ATTASYLLLL
VLHYFVAKKK YAYDEVAIST FVKVIALVVV YTGLMTVFVG SIWIRWSLGI
AVLVVYAYIF RKELTVALNT FREKRSK

Fig. 3 cont. CPS20

MVYIIAEIGC NHNGDVHLAR KMVEVAVDCG VDAVKFQTFK ADLLISKYAP KAEYQKITTG
ESDSQLEMTR RLELSFEEYL DLRDYCLEKG VDVFSTPFDE ESLDFLISTD
MPVYKIPSGE ITNLPYLEKI GRQAKKVILS TGMAVMDEIH QAVKILQENG TTDISILHCT
TEYPTPYPAL NLNVLHTLKK EFPNLTIGYS DHSVGSEVPI AAAAMGAELI
EKHFTLDNEM EGPDHKASAT PDILAALVKG VRIVEQSLGK FEKEPEEVEV RNKIVARKSI
VAKKAIAKGE VFTEENITVK RPGNGISPME WYKVLGQVSE QDFEEDQNIC
HSAFENQM

Fig. 3 cont. CPS2P

MKKICFVTGS RAEYGIMRRL LSYLQDDPEM ELDLVVTAMH LEEKYGMTVK DIEADKRRIV
KRIPLHLTDT SKQTIVKSLA TLTEQLTVLF EEVQYDLVLI LGDRYEMLPV
ANAALLYNIP ICHIHGGEKT MGNFDESIRH AITKMSHLHL TSTDEFRNRV IQLGENPTMY

Fig. 3 cont. CPS2Q

MELGIDFAED YYVVLFHPVT LEDNTAEEQT QALLDALKED GSQCLIIGSN SDTHADKIME
LMHEFVKQDS DSYIFTSLPT RYYHSLVKHS QGLIGNSSSG LIEVPSLQVP
TLNIGNRQFG RLSGPSVVHV GTSKEAIVGG LGQLRDVIDF TNPFEQPDSA LQGYRAIKEF
LSVQASTMKE FYDR

Fig. 3 cont. CPS2R

MKKVAFLGAG TFSDGVLPWL DRTRYELIGY FEDKPISDYR GYPVFGPLQD VLTYLDDGKV
DAVFVTIGDN VKRKEIFDLL AKDHYDALFN IISEQANIFS PDSIKGRGVF
IGFSSFVGAD SYVYDNCIIN TGAIVEHHTT VEAHCNITPG VTINGLCRIG ESTYIGSGST
VIQCIEIAPY TTLGAGTVVL KSLTESGTYV GVPARKIK

Fig. 3 cont. CPS2S

MEPICLIPAR SGSKGLPNKN MLFLDGVPMI FHTIRAAIES GCFKKENIYV STDSEVYKEI
CETTGVQVLM RPADLATDFT TSFQLNEHFL QDFSDDQVFV LLQVTSPLRS
GKHVKEAMEL YGKGQADHVV SFTKVDKSPT LFSTLDENGF AKDIAGLGGS YRRQDEKTLY
YPNGAIYISS KQAYLADKTY FSEKTAAYVM TKEDSIDVDD HFDFTGVIGR
IYFDYQRREQ QNKPFYKREL KRLCEQRVHD SLVIGDSRLL ALLLDGFDNI SIGGMTASTA
LENQGLFLAT PIKKVLLSLG VNDLITDYPL HMIEDTIRQL MESLVSKAEQ
VFVTTIAYTL FRDSVSNEEI VQLNDVIVQS ASELGISVID LNEVVEKEAM LDYQYTNDGL
HFNQIGQERV NQLILTSLTR

Fig. 3 cont. CPS2T

ATCGCCAAAC GAAATTGGCA TTATTTGATA TGATAGCAGT TGCAATTTCT GCAATCTTAA CAAGTCATAT
ACCAAATGCT GATTTAAATC GTTCTGGAAT TTTTATCATA
ATGATGGTTC ATTATTTTGC ATTTTTTATA TCTCGTATGC CAGTTGAATT TGAGTATAGA GGTAATCTGA
TAGAGTTTGA AAAAACATTT AACTATAGTA TAATATTTGC
AATTTTTCTT ACGGCAGTAT CATTTTTGTT GGAGAATAAT TTCGCACTTT CAAGACGTGG TGCCGTGTAT
TTCACATTAA TAAACTTCGT TTTGGTATAC CTATTTAACG
TAATTATTAA GCAGTTTAAG GATAGCTTTC TATTTTCGAC AATCTATCAA AAAAAGACGA TTCTAATTAC
AACGGCTGAA CGATGGGAAA ATATGCAAGT TTTATTTGAA
TCACATAAAC AAATTCAAAA AAATCTTGTT GCATTGGTAG TTTTAGGTAC AGAAATAGAT AAAATTAATT
TATCATTACC GCTCTATTAT TCTGTGGAAG AAGCTATAGA
GTTTTCAACA AGGGAAGTGG TCGACCACGT CTTTATAAAT CTACCAAGTG AGTTTTTAGA CGTAAAGCAA
TTCGTTTCAG ATTTTGAGTT GTTAGGTATT GATGTAAGCG
TTGATATTAA TTCATTCGGT TTTACTGCGT TGAAAAACAA AAAAATCCAA CTGCTAGGTG ACCATAGCAT
TGTAACTTTT TCCACAAATT TTTATAAGCC TAGTCATATC
ATGATGAAAC GACTTTTGGA TATACTCGGA GCGGTAGTCG GGTTAATTAT TTGTGGTATA GTTTCTATTT
TGTTAGTTCC AATTATTCGT AGAGATGGTG GACCGGCTAT
TTTTGCTCAG AAACGAGTTG GACAGAATGG ACGCATATTT ACATTCTACA AGTTTCGATC GATGTATGTT
GATGCTGAGG AGCGCAAAAA AGACTTGCTC AGCCAAAACC
AGATGCAAGG GTGGGTATGT TTTAAAATGG GAAAAACGAT CCTAGAATTA CTCCAATTGG ACATTTCATA
CGCAAAAACA AGTTTAGACG AGTTACCACA GTTTTATAAT
GTTTTAATTG GCGATATGAG TCTAGTTGGT ACACGTCCAC CTACAGTTGA TGAATTTGAA AAATATACTC
CTGGTCAAAA GAGACGATTG AGTTTTAAAC CAGGGATTAC
AGGTCTCTGG CAGGTTAGTG GTCGTAGTAA TATCACAGAC TTCGACGACG TAGTTCGGTT GGACTTAGCA
TACATTGATA ATTGGACTAT CTGGTCAGAT ATTAAAATTT
TATTAAAGAC AGTGAAAGTT GTATTGTTGA GAGAGGGAAG TAAGTAAAAG TATATGAAAG TTTGTTTGGT
CGGTTCTTCA GGGGGACATT TGACTCACTT GTATTTGTTA
AAACCGTTTT GGAAGGAAGA AGAACGTTTT TGGGTAACAT TTGATAAAGA GGATGCAAGA AGTCTTTTGA
AGAATGAAAA AATGTATCCA TGTTACTTTC CAACAAATCG
CAATCTCATT AATTTAGTGA AAAATACTTT CTTAGCTTTC AAAATTTTAC GTGATGAGAA ACCAGATGTT
ATTATTTCAT CTGGTGCGGC CGTTGCTGTC CCCTTCTTTT
ACATCGGAAA ACTATTTGGA GCAAAGACGA TTTATATTGA AGTATTTGAT CGAGTTAATA AATCTACATT
AACTGGAAAA CTAGTTTATC CCGTAACAGA TATTTTTATT
GTTCAGTGGG AAGAAATGAA GAAGGTATAT CCTAAATCTA TTAACTTGGG GAGTATTTTT TAATGATTTT
TGTAACAGTA GGAACTCATG AACAACAGTT TAATCGATTG
ATAAAAGAGA TTGATTTATT GAAAAAAAAT GGAAGTATAA CCGACGAAAT ATTTATTCAA ACAGGATATT
CTGACTATAT TCCAGAATAT TGCAAGTATA AAAAATTTCT
CAGTTACAAA GAAATGGAAC AATATATTAA CAAATCAGAA GTAGTTATTT GCCACGGAGG CCCCGCTACT
TTTATGAATT CATTATCCAA AGGAAAAAAA CAATTATTGT
TTCCTAGACA AAAAAAGTAT GGTGAACATG TAAATGATCA TCAAGTAGAG TTTGTAAGAA GAATTTTACA
AGATAATAAT ATTTTATTTA TAGAAAATAT AGATGATTTG
TTTGAAAAAA TTATTGAAGT TTCTAAGCAA ACTAACTTTA CATCAAATAA TAATTTTTTT TGTGAAAGAT
TAAAACAAAT AGTTGAAAAA TTTAATGAGG ATCAAGAAAA
TGAATAATAA AAAAGATGCA TATTTGATAA TGGCTTATCA TAATTTTTCT CAGATTTTAC TGGAGAGGGA
TACAGATATT ATCATCTTCT CTCAGGAGAA TGCACACCAT
TAGTTCCTTC AGAATACCTG TATAATTATT TTAAATATTC TCAGGATTTA TATGTTGAAT TTACAAAAGA
TGAGCAAAAA TATAAAGAAA ATAGGATATA TGAACGAGTT
AAATGTTACA GATTATTTCC TAATATATCA GAAAAAACTA TTGATAATGT ACTGTTTAGA ATTTTATTAA
GAATGTATCG AGCTTTTGAA TACTATTTAC AAAGATTGTT
GTTTATTGAT AGAATAAAAA ACATGGTCTA AGAATAAGAT TTGGTTCTAA TTGGGTTTCG CTTCCACATG
ATTTTGTGGC AATTCTTTTA TCAAATGAAA ACGAAACAGC
TTATTTATTT AAGTAATCTA AATGTCCAGA TGAACTATTT ATACAGACAA TTATAGAAAA ATATGAATTT
TCAAATAGAT TATCTAAATA TGGAAATTTA AGATATATAA
AGTGGAAAAA ATCAACATCT TCTCCTATTG TCTTTACAGA TGATTCTATT GATGAATTGC TAAATGCAAG
AAATTTAGGT TTTTTATTTG CTAGAAAGTT AAAAATAGAA
AATAAATCTA AATTTAAAGA AATTATTACT AAAAAATAAA ATAGTTGATT TTGTGAGAGT AATGTATGTT
TAAATTATTT AAATATGACC CGGAATATTT TATTTTTAAG
TACTTCTGGT TGATTATTTT TATTCCAGAG CAAAAGTATG TATTTTTATT AATTTTTATG AATTTAATTT
TATTTCATAT AAAATTTTTG AAAACTAAGC TAATATTAAA
AAATGAAATT TTATTGTTTT TATTATGGTC TATATTATGT TTTGTTTCAG TAGTCACAAG TATGTTTGTT
GAAATAAATT TTGAAAGATT ATTTGCAGAT TTTACTGCTC
CCATAATTTG GATTATTGCA ATAATGTATT ATAATTTGTA TTCATTTATA AATATTGATT ATAAAAAATT
AAAAAATAGT ATCTTTTTTA GTTTTTTAGT TTTATTAGGT
ATATCTGCAT TGTATATTAT TCAAAATGGG AAAGATATTG TATTTTTAGA CAGACACCTT ATAGGACTAG
ACTATCTTAT AACAGGCGTC AAAACAAGGT TGGTTGGCTT
TATGAACTAT CCTACGTTAA ATACCACTAC AATTATAGTT TCAATTCCGT TAATCTTTGC ACTTATAAAA
AATAAAATGC AACAATTTTT TTTCTTGTGT CTTGCTTTTA

Fig. 4

TACCGATCTA TTTAAGTGGA TCGAGAATTG GTAGTTTATC GCTAGCAATA TTAATTATAT GCTTGTTATG
GAGATATATA GGTGGAAAAT TTGCTTGGAT AAAAAAGCTA
ATAGTAATAT TTGTAATACT ACTTATTATT TTAAATACTG AATTGCTTTA CCATGAAATT TTGGCTGTTT
ATAATTCTAG AGAATCAAGT AACGAAGCTA GATTTATTAT
TTATCAAGGA AGTATTGATA AAGTATTAGA AAACAATATT TTATTTGGAT ATGGAATATC CGAATATTCA
GTTACGGGAA CTTGGCTCGG AAGTCATTCA GGCTATATAT
CATTTTTTTA TAAATCAGGA ATAGTTGGGT TGATTTTACT GATGTTTTCT TTTTTTTATG TTATAAAAAA
AAGTTATGGA GTTAATGGGG AAACAGCACT ATTTTATTTT
ACATCATTAG CCATATTTTT CATATATGAA ACAATAGATC CGATTATTAT TATATTAGTA CTATTCTTTT
CTTCAATAGG TATTTGGAAT AATATAAATT TTAAAAAGGA
TATGGAGACA AAAAATGAAT GATTTAATTT CAGTTATTGT ACCAATTTAT AATGTCCAAG ATTATCTTGA
TAAATGTATT AACAGTATTA TTAACCAAAC ATATACTAAT
TTAGAGGTTA TTCTCGTAAA TGATGGAAGT ACTGATGATT CTGAGAAAAT TTGCTTAAAC TATATGAAGA
ACGATGGAAG AATTAAATAT TACAAGAAAA TTAATGGCGG
TCTAGCAGAT GCTCGAAATT TCGGACTAGA ACATGCAACA GGTAAATATA TTGCTTTTGT CGATTCTGAT
GACTATATAG AAGTTGCAAT GTTCGAGAGA ATGCATGATA
ATATAACTGA GTATAATGCC GATATAGCAG AGATAGATTT TTGTTTAGTA GACGAAAACG GGTATACAAA
GAAAAAAAGA AATAGTAATT TTCATGTCTT AACGAGAGAA
GAGACTGTAA AAGAATTTTT GTCAGGATCT AATATAGAAA ATAATGTTTG GTGCAAGCTT TATTCACGAG
ATATTATAAA AGATATAAAA TTCCAAATTA ATAATAGAAG
TATTGGTGAG GATTTGCTTT TTAATTTGGA GGTCTTGAAC AATGTAACAC GTGTAGTAGT TGATACTAGA
GAATATTATT ATAATTATGT CATTCGTAAC AGTTCGCTTA
TTAATCAGAA ATTCTCTATA AATAATATTG ATTTAGTCAC AAGATTGGAG AATTACCCCT TTAAGTTAAA
AAGAGAGTTT AGTCATTATT TTGATGCAAA AGTTATTAAA
GAGAAGGTTA AATGTTTAAA CAAAATGTAT TCAACAGATT GTTTGGATAA TGAGTTCTTG CCAATATTAG
AGTCTTATCG AAAAGAAATA CGTAGATATC CATTTATTAA
AGCGAAAAGA TATTTATCAA GAAAGCATTT AGTTACGTTG TATTTGATGA AATTTTCGCC TAAACTATAT
GTAATGTTAT ATAAGAAATT TCAAAAGCAG TAGAGGTAAA
AATGGATAAA ATTAGTGTTA TTGTTCCAGT TTATAATGTA GATAAATATT TAAGTAGTTG TATAGAAAGC
ATTATTAATC AAAATTATAA AAATATAGAA ATATTATTGA
TAGATGATGG CTCTGTAGAT GATTCTGCTA AAATATGCAA GGAATATGCA GAAAAAGATA AAAGAGTAAA
AATTTTTTTC ACTAATCATA GTGGAGTATC AAATGCTAGA
AATCATGGAA TAAAGCGGAG TACAGCTGAA TATATTATGT TTGTTGACTC TGATGATGTT GTTGATAGTA
GATTAGTAGA AAAATTATAT TTTAATATTA TAAAAAGTAG
AAGTGATTTA TCTGGTTGTT TGTACGCTAC TTTTTCAGAA AATATAAATA ATTTTGAAGT GAATAATCCA
AATATTGATT TTGAAGCAAT TAATACCGTG CAGGACATGG
GAGAAAAAAA TTTTATGAAT TTGTATATAA ATAATATTTT TTCTACTCCT GTTTGTAAAC TATATAAGAA
AAGATACATA ACAGATCTTT TTCAAGAGAA TCAATGGTTA
GGAGAAGATT TACTTTTTAA TCTGCATTAT TTAAAGAATA TAGATAGAGT TAGTTATTTG ACTGAACATC
TTTATTTTTA TAGGAGAGGT ATACTAAGTA CAGTAAATTC
TTTTAAAGAA GGTGTGTTTT TGCAATTGGA AAATTTGCAA AAACAAGTGA TAGTATTGTT TAAGCAAATA
TATGGTGAGG ATTTTGACGT ATCAATTGTT AAAGATACTA
TACGTTGGCA AGTATTTTAT TATAGCTTAC TAATGTTTAA ATACGGAAAA CAGTCTATTT TTGACAAATT
TTTAATTTTT AGAAATCTTT ATAAAAAATA TTATTTTAAC
TTGTTAAAAG TATCTAACAA AAATTCTTTG TCTAAAAATT TTTGTATAAG AATTGTTTCG AACAAAGTTT
TTAAAAAAAT ATTATGGTTA TAATAGGAAG ATATCATGGA
TACTATTAGT AAAATTTCTA TAATTGTACC TATATATAAT GTAGAAAAAT ATTTATCTAA ATGTATAGAT
AGCATTGTAA ATCAGACCTA CAAACATATA GAGATTCTTC
TGGTGAATGA CGGTAGTACG GATAATTCGG AAGAAATTTG TTTAGCATAT GCGAAGAAAG ATAGTCGCAT
TCGTTATTTT AAAAAAGAGA ACGGCGGGCT ATCAGATGCC
CGTAATTATG GCATAAGTCG CGCCAAGGGT GACTACTTAG CTTTTATAGA CTCAGATGAT TTTATTCATT
CGGAGTTCAT CCAACGTTTA CACGAAGCAA TTGAGAGAGA
GAATGCCCTT GTGGCAGTTG CTGGTTATGA TAGGGTAGAT GCTTCGGGGC ATTTCTTAAC AGCAGAGCCG
CTTCCTACAA ATCAGGCTGT TCTGAGCGGC AGGAATGTTT
GTAAAAAGCT GCTAGAGGCG GATGGTCATC GCTTTGTGGT GGCCTGTAAT AAACTCTATA AAAAAGAACT
ATTTGAAGAT TTTCGATTTG AAAAGGGTAA GATTCATGAA
GATGAATACT TCACTTATCG CTTGCTCTAT GAGTTAGAAA AAGTTGCAAT AGTTAAGGAG TGCTTGTACT
ATTATGTTGA CCGAGAAAAT AGTATCACAA CTTCTAGCAT
GACTGACCAT CGCTTCCATT GCCTACTGGA ATTTCAAAAT GAACGAATGG ACTTCTATGA AAGTAGAGGA
GATAAAGAGC TCTTACTAGA GTGTTATCGT TCATTTTTAG
CCTTTGCTGT TTTGTTTTTA GGCAAATATA ATCATTGGTT GAGCAAACAG CAAAAGAAGC TT

Fig. 4 cont.

RQTKLALFDM IAVAISAILT SHIPNADLNR SGIFIIMMVH YFAFFISRMP VEFEYRGNLI
EFEKTFNYSI IFAIFLTAVS FLLENNFALS RRGAVYFTLI NFVLVYLFNV
IIKQFKDSFL FSTIYQKKTI LITTAERWEN MQVLFESHKQ IQKNLVALVV LGTEIDKINL
SLPLYYSVEE AIEFSTREVV DHVFINLPSE FLDVKQFVSD FELLGIDVSV
DINSFGFTAL KNKKIQLLGD HSIVTFSTNF YKPSHIMMKR LLDILGAVVG LIICGIVSIL
LVPIIRRDGG PAIFAQKRVG QNGRIFTFYK FRSMYVDAEE RKKDLLSQNQ
MQGWVCFKMG KTILELLQLD ISYAKTSLDE LPQFYNVLIG DMSLVGTRPP TVDEFEKYTP
GQKRRLSFKP GITGLWQVSG RSNITDFDDV VRLDLAYIDN WTIWSDIKIL
LKTVKVVLLR EGSK

Fig. 4 cont. CPS1E

MKVCLVGSSG GHLTHLYLLK PFWKEEERFW VTFDKEDARS LLKNEKMYPC YFPTNRNLIN
LVKNTFLAFK ILRDEKPDVI ISSGAAVAVP FFYIGKLFGA KTIYIEVFDR
VNKSTLTGKL VYPVTDIFIV QWEEMKKVYP KSINLGSIF

Fig. 4 cont. CPS1F

```
MIFVTVGTHE QQFNRLIKEI DLLKKNGSIT DEIFIQTGYS DYIPEYCKYK KFLSYKEMEQ
YINKSEVVIC HGGPATFMNS LSKGKKQLLF PRQKKYGEHV NDHQVEFVRR
ILQDNNILFI ENIDDLFEKI IEVSKQTNFT SNNNFFCERL KQIVEKFNED QENE
```

Fig. 4 cont. CPS1G

MFKLFKYDPE YFIFKYFWLI IFIPEQKYVF LLIFMNLILF HIKFLKTKLI LKNEILLFLL
WSILCFVSVV TSMFVEINFE RLFADFTAPI IWIIAIMYYN LYSFINIDYK
KLKNSIFFSF LVLLGISALY IIQNGKDIVF LDRHLIGLDY LITGVKTRLV GFMNYPTLNT
TTIIVSIPLI FALIKNKMQQ FFFLCLAFIP IYLSGSRIGS LSPLAILIIC
LLWRYIGGKF AWIKKLIVIF VILLIILNTE LLYHEILAVY NSRESSNEAR FIIYQGSIDK
VLENNILFGY GISEYSVTGT WLGSHSGYIS FFYKSGIVGL ILLMFSFFYV
IKKSYGVNGE TALFYFTSLA IFFIYETIDP IIIILVLFFS SIGIWNNINF KKDMETKNE

Fig. 4 cont. CPS1H

MNDLISVIVP IYNVQDYLDK CINSIINQTY TNLEVILVND GSTDDSEKIC LNYMKNDGRI
KYYKKINGGL ADARNFGLEH ATGKYIAFVD SDDYIEVAMF ERMHDNITEY
NADIAEIDFC LVDENGYTKK KRNSNFHVLT REETVKEFLS GSNIENNVWC KLYSRDIIKD
IKFQINNRSI GEDLLFNLEV LNNVTRVVVD TREYYYNYVI RNSSLINQKF
SINNIDLVTR LENYPFKLKR EFSHYFDAKV IKEKVKCLNK MYSTDCLDNE FLPILESYRK
EIRRYPFIKA KRYLSRKHLV TLYLMKFSPK LYVMLYKKFQ KQ

Fig. 4 cont. CPS1I

MDKISVIVPV YNVDKYLSSC IESIINQNYK NIEILLIDDG SVDDSAKICK EYEKDKRVKI
FFTNHSGVSN ARNHGIKRST AEYIMFVDSD DVVDSRLVEK LYFNIIKSRS
DLSGCLYATF SENINNFEVN NPNIDFEAIN TVQDMGEKNF MNLXXNNIFS TPVCXLYQKR
YITDLFQENQ WLGEDLLFNL HYLKNIDRVS YLTEHLYFYR RGILSTVNSF
KEGVFLQLEN LQKQVIVLFK QIYGEDFDVS IVKDTIRWQV FYYSLLMFKY GKQSIFDKFL
IFRNLYKKYY FNLLKVSNKN SLSKNFCIRI VSNKVFKKIL WL

Fig. 4 cont. CPS1J

MDTISKISII VPIYNVEKYL SKCIDSIVNQ TYKHIEILLV NDGSTDNSEE ICLAYAKKDS
RIRYFKKENG GLSDARNYGI SRAKGDYLAF IDSDDFIHSE FIQRLHEAIE
RENALVAVAG YDRVDASGHF LTAEPLPTNQ AVLSGRNVCK KLLEADGHRF VVACNKLYKK
ELFEDFRFEK GKIHEDEYFT YRLLYELEKV AIVKECLYYY VDRENSITTS
SMTDHRFHCL LEFQNERMDF YESRGDKELL LECYRSFLAF AVLFLGKYNH WLSKQQKK

Fig. 4 cont. CPS1K

AAGCTTATCG TCAAGGTGTT CGCTATATCG TGGCGACATC TCATAGACGA AAAGGGATGT
TTGAAACACC AGAAAAAGTT ATCATGACTA ACTTTCTTCA ATTTAAAGAC
GCAGTAGCAG AAGTTTATCC TGAAATACGA TTGTGCTATG GTGCTGAATT GTATTATAGT
AAAGATATAT TAAGCAAACT TGAAAAAAAG AAAGTACCCA CACTTAATGG
CTCGCGCTAT ATTCTTTTGG AGTTCAGTAG TGATACTCCT TGGAAAGAGA TTCAAGAAGC
AGTGAACGAA GTGACGCTAC TTGGGCTAAC TCCCGTACTT GCCCATATAG
AACGATATGA CGCCCTAGCG TTTCATGCAG AGAGAGTAGA AGAGTTAATT GACAAGGGAT
GCTATACTCA GGTAAATAGT AATCATGTGC TGAAGCCCAC TTTAATTGGT
GATCGAGCAA AAGAATTTAA AAAACGTACT CGGTATTTTT TAGAGCAGGA TTTAGTACAT
TGTGTTGCTA GCGATATGCA TAATTTATCT AGTAGACCTC CGTTTATGAG
GGAGGCTTAT AAGTTGCTAA CAGAGGAATT TGGCAAAGAT AAAGCGAAAG CGTTGCTAAA
AAAGAATCCT CTTATGCTAT TAAAAAACCA GGCGATTTAA ACTGGTTACT
CTAGATTGTG GAGAGAAAAA TGGATTTAGG AACTGTTACT GATAAACTGT TAGAACGCAA
CAGTAAACGA TTGATACTCG TGTGCATGGA TACGTGTCTT CTTATAGTTT
CCATGATTTT GAGCAGACTG TTTTTGGATG TTATTATTGA CATACCAGAT GAACGCTTCA
TTCTTGCAGT TTTATTCGTA TCAATTTTAT ATTTGATTCT ATCGTTTAGA
TTAAAAGTCT TTTCATTAAT TACGCGTTAC ACAGGGTATC AGAGTTATGT AAAAATAGGA
CTTAGTTTAA TATCTGCGCA TTCATTGTTT TTAATTATCT CAATGGTGTT
GTGGCAGGCT TTTAGTTATC GTTTCATCTT AGTATCCTTA TTTTTGTCGT ATGTAATGCT
CATTACTCCG AGGATTGTTT GGAAAGTCTT ACATGAGACG AGAAAAAATG
CTATCCGTAA GAAGGATAGC CCACTAAGAA TCTTAGTAGT AGGTGCTGGA GATGGTGGTA
ATATTTTTAT CAATACTGTC AAAGATCGAA AATTGAATTT TGAAATTGTC
GGTATCGTTG ATCGTGATCC AAATAAACTT GGAACATTTA TCCGTACGGC TAAAGTTTTA
GGAAACCGTA ATGATATTCC ACGACTGGTA GAGGAATTAG CTGTTGACCA
AGTGACGATT GCCATCCCTT CTTTAAATGG TAAGGAGCGA GAGAAGATTG TTGAAATCTG
TAACACTACA GGAGTGACCG TCAATAATAT GCCGAGTATT GAAGACATTA
TGGCGGGGAA CATGTCTGTC AGTGCCTTTC AGGAAATTGA CGTAGCAGAC CTTCTTGGTC
GACCAGAGGT TGTTTTGGAT CAGGATGAAT TGAATCAGTT TTTCCAAGGG
AAAACAATCC TTGTCACAGG AGCAGGTGGC TCTATCGGTT CAGAGCTATG TCGTCAAATT
GCTAAGTTTA CGCCTAAACG CTTGTTGTTG CTTGGACATG GAGAAAATTC
AATCTATCTC ATTCATCGAG AGTTACTGGA AAAGTACCAA GGTAAGATTG AGTTGGTCCC
TCTCATTGCA GATATTCAAG ATAGAGAATT GATTTTTAGC ATAATGGCTG
AATATCAACC CGATGTTGTT TATCATGCTG CAGCACATAA GCATGTTCCT TTGATGGAAT
ATAATCCACA TGAAGCAGTG AAGAATAATA TTTTTGGAAC GAAGAATGTG
GCTGAGGCGG CTAAAACTGC AAAGGTTGCC AAATTTGTTA TGGTTTCAAC AGATAAAGCT
GTTAATCCAC CAAATGTCAT GGGAGCGACT AAACGTGTTG CAGAAATGAT
TGTTACAGGT TTAAACGAGC CAGGTCAGAC TCAATTTGCG GCAGTCCGGT TTGGGAATGT
TCTAGGTAGT CGTGGAAGTG TTGTTCCGCT ATTCAAAGAG CAAATTAGAA
AAGGTGGACC TGTTACGGTT ACCGACTTTA GGATGACTCG TTATTTCATG ACGATTCCTG
AGGCAAGTCG TTTGGTTATC CAAGCTGGAC ATTTGGCAAA AGGTGGAGAA
ATATTTGTCT TGGATATGGG CGAGCCAGTA CAAATCCTGG AATTGGCAAG AAAAGTTATC
TTGTTAAGTG GACACACAGA GGAAGAAATC GGGATTGTAG AATCTGGAAT
CAGACCAGGC GAGAAACTCT ACGAGGAATT ATTATCAACA GAAGAACGTG TCAGCGAACA
GATTCATGAA AAAATATTTG TGGGTCGCGT TACAAATAAG CAGTCGGACA
TTGTCAATTC ATTTATCAAT GGATTACTCC AAAAAGATAG AAATGAATTA AAAAATATGT
TGATTGAATT TGCAAAACAA GAATAAGAAA GTAAAAAATA TTTTTACTTT
CCTAGAGTTT AAACGATGTT TAAGTTCTAG GAAGGTTAGA ATACCTAATT AACAACAATA
TTACTATTTA TTAAGAGTCA GATAATAGCA ACTAAGTGCT ACAAACTATC
TTTATAATAA GTATATTTGG TCAAAAGGGA GATGTGAAAT GTATCCAATT TGTAAACGTA
TTTTAGCAAT TATTATCTCA GGGATTGCTA TTGTTGTTCT GAGTCCAATT
TTATTATTGA TTGCATTGGC AATTAAATTA GATTCTAAAG GTCCGGTATT ATTTAAACAA
AAGCGGGTTG GTAAAAACAA GTCATACTTT ATGATTTATA AATTCCGTTC
TATGTACGTT GACGCACCAA GTGATATGCC GACTCATCTA TTAAAGGATC CTAAGGCGAT
GATTACCAAG GTGGGCGCGT TTCTCAGAAA AACAAGTTTA GATGAACTGC
CACAGCTTTT TAATATTTTT AAAGGTGAAA TGGCGATTGT TGGTCCACGC CCAGCCTTAT
GGAATCAATA TGACTTAATT GAAGAGCGAG ATAAATATGG TGCAAATGAT
ATTCGTCCTG GACTAACCGG TTGGGCTCAA ATTAATGGTC GTGATGAATT GGAAATTGAT
GAAAAGTCAA AATTAGATGG ATATTATGTT CAAAATATGA GTCTAGGTTT
GGATATTAAA TGTTTCTTAG GTACATTCCT CAGTGTAGCC AGAAGCGAAG GTGTTGTTGA
AGGTGGAACA GGGCAGAAAG GAAAAGGATG AAATTTTCAG TATTAATGTC
GGTCTATGAG AAAGAAAAAC CAGAGTTTCT TAGGGAATCT TTGGAAAGCA TCCTTGTCAA
TCAAACAATG ATTCCAACGG AGGTTGTCTT GGTAGAGGAT GGGCCACTCA
ATCAGAGCTT ATATAGTATT TTAGAAGAAT TTAAAAGTCG ATTTTCATTT TTTAAAACGA
TAGCCTTGGA AAAGAATTCG GGTTTAGGAA TTGCACTGAA TGAAGGTTTG
AAACATTGTA ATTATGAGTG GGTTTGCACG AAATGGATTC TGATGATGTT GCATATACAT
ACACGTTTTG AAAAGCAAGT TAACTTTATA AAACAAAACC CGACTATAGA

DNA Serotype 9

TATTGAGATA GATGAGTTCT TAAATTCTAC TAGTGAAATA GTTTCTCATA AAAATGTTCC
AACCCAGCAC GATGAAATAT TAAAGATGGC AAGGCGGGAG AAATCCATGT
GCCACATGAC TGTAATGTTT AAAAAGAAAA GTGTCGAGAG AGCAGGGGGG TATCAAACAC
TTCCGTACGT AGAAGATTAT TTCCTTTGGG TGCGCATGAT TGCTTCAGGA
TCGAAATTTG CAAACATTGA TGAAACACTA GTTCTTGCAC GTGTTGGAAA TGGGATGTTC
AATAGGAGGG GGAACAGAGA ACAAATTAAC AGTTGGACAT TACTAATTGA
ATTTATGTTA GCTCAAGGAA TTGTTACACC ACTAGATGTA TTTATTAATC AAATTTACAT
TAGGGTCTTT GTTTATATGC CAACTTGGAT AAAGAAACTC ATTTATGGAA
AAATCTTAAG GAAATAGTAT GATTACAGTA TTGATGGCTA CATATAATGG AAGCCCATTT
ATAATAAAAC AGTTAGATTC AATTCGAAAT CAAAGTGTAT CAGCAGACAA
AGTTATTATT TGGGATGATT GCTCGACAGA TGATACAATA AAAATAATAA AAGATTATAT
AAAAAAATAT TCTTTGGATT CATGGGTTGT CTCTCAAAAT AAATCTAATC
AGGGGCATTA TCAAACATTT ATAAATTTGA CAAAGTTAGT TCAGGAAGGA ATAGTCTTTT
TTTCAGATCA AGATGATATT TGGGACTGTC ATAAAATTGA GACAATGCTT
CCAATCTTTG ACAGAGAAAA TGTATCAATG GTGTTTTGCA AATCCAGATT GATTGATGAA
AACGGAAATA TTATCAGTAG CCCAGATACT TCGGATAGAA TCAATACGTA
CTCTCTAGA

Fig. 5 cont.

AYRQGVRYIV ATSHRRKGMF ETPEKVIMTN FLQFKDAVAE VYPEIRLCYG AELYYSKDIL
SKLEKKKVPT LNGSRYILLE FSSDTPWKEI QEAVNEVTLL GLTPVLAHIE
RYDALAFHAE RVEELIDKGC YTQVNSNHVL KPTLIGDRAK EFKKRTRYFL EQDLVHCVAS
DMHNLSSRPP FMREAYKLLT EEFGKDKAKA LLKKNPLMLL KNQAI

Fig. 5 cont. CPS9D

MDLGTVTDKL LERNSKRLIL VCMDTCLLIV SMILSRLFLD VIIDIPDERF ILAVLFVSIL
YLILSFRLKV FSLITRYTGY QSYVKIGLSL ISAHSLFLII SMVLWQAFSY
RFILVSLFLS YVMLITPRIV WKVLHETRKN AIRKKDSPLR ILVVGAGDGG NIFINTVKDR
KLNFEIVGIV DRDPNKLGTF IRTAKVLGNR NDIPRLVEEL AVDQVTIAIP
SLNGKEREKI VEICNTTGVT VNNMPSIEDI MAGNMSVSAF QEIDVADLLG RPEVVLDQDE
LNQFFQGKTI LVTGAGGSIG SELCRQIAKF TPKRLLLLGH GENSIYLIHR
ELLEKYQGKI ELVPLIADIQ DRELIFSIMA EYQPDVVYHA AAHKHVPLME YNPHEAVKNN
IFGTKNVAEA AKTAKVAKFV MVSTDKAVNP PNVMGATKRV AEMIVTGLNE
PGQTQFAAVR FGNVLGSRGS VVPLFKEQIR KGGPVTVTDF RMTRYFMTIP EASRLVIQAG
HLAKGGEIFV LDMGEPVQIL ELARKVILLS GHTEEEIGIV ESGIRPGEKL
YEELLSTEER VSEQIHEKIF VGRVTNKQSD IVNSFINGLL QKDRNELKNM LIEFAKQE

Fig. 5 cont. CPS9E

MYPICKRILA IIISGIAIVV LSPILLLIAL AIKLDSKGPV LFKQKRVGKN KSYFMIYKFR
SMYVDAPSDM PTHLLKDPKA MITKVGAFLR KTSLDELPQL FNIFKGEMAI
VGPRPALWNQ YDLIEERDKY GANDIRPGLT GWAQINGRDE LEIDEKSKLD GYYVQNMSLG
LDIKCFLGTF LSVARSEGVV EGGTGQKGKG

Fig. 5 cont. CPS9F

MKFSVLMSVY EKEKPEFLRE SLESILVNQT MIPTEVVLVE DGPLNQSLYS ILEEFKSRFS
FFKTIALEKN SGLGIALNEG LKHCNYEWVC TKWILMMLHI HTRFEKQVNF
IKQNPTIDIE IDEFLNSTSE IVSHKNVPTQ HDEILKMARR EKSMCHMTVM FKKKSVERAG
GYQTLPYVED YFLWVRMIAS GSKFANIDET LVLARVGNGM FNRRGNREQI
NSWTLLIEFM LAQGIVTPLD VFINQIYIRV FVYMPTWIKK LIYGKILRK

Fig. 5 cont. CPS9G

MITVLMATYN GSPFIIKQLD SIRNQSVSAD KVIIWDDCST DDTIKIIKDY IKKYSLDSWV
VSQNKSNQGH YQTFINLTKL VQEGIVFFSD QDDIWDCHKI ETMLPIFDRE
NVSMVFCKSR LIDENGNIIS SPDTSDRINT YSL

Fig. 5 cont. CPS9H

CTGCAGCACA TAAGCATGTT CCATTGATGG AATATAATCC ACATGAAGCA GTGAAGAATA
ATATTTTTGG AACGAAGAAT GTGGCTGAGG CGGCTAAAAC TGCAAAGGTT
GCCAAATTTG TTATGGTTTC AACAGATAAA GCTGTTAATC CGCCAAATGT CATGGGAGCG
ACTAAACGTG TTGCAGAAAT GATTGTAACA GGTTTAAACG AGCCAGGTCA
GACTCAATTT GCGGCAGTCC GTTTTGGGAA TGTTCTAGGT AGTCGTGGAA GTGTTGTTCC
GCTATTCAAA GAGCAAATTA GAAAAGGTGG ACCTGTTACG GTTACCGACT
TTAGGATGAC TCGTTATTTC ATGACGATTC CTGAGGCAAG TCGTTTGGTT ATCCAAGCTG
GACATTTGGC AAAAGGTGGA GAAATCTTTG TCTTGGATAT GGGTGAGCCA
GTACAAATCC TGGAATTGGC AAGAAAAGTT ATCTTGTTAA GCGGACATAC AGAGGAAGAA
ATCGGGATTG TAGAATCTGG AATCAGACCA GGCGAGAAAC TCTACGAGGA
ATTGTTATCA ACAGAAGAAC GTGTCAGCGA ACAGATTCAT GAAAAAATAT TTGTGGGTCG
CGTTACAAAT AAGCAGTCGG ACATTGTCAA TTCATTTATC AATGGATTAC
TCCAAAAAGA TAGAAATGAA TTAAAAGATA TGTTGATTGA ATTTGCAAAA CAAGAATAAG
AAAGTAAAAA ATATTTTTAC TTTCCTAGAG TTTAAACGAT GTTTAAGTTC
TAGGAAGGTT GGAATTGCTT TCGTGGAGGT GATAGATAGA AACCTATATA TTTGTAGAAG
AAAGGATATT AAACTAAAGG TGAATCGGAA CATAAAGTTT AGATAGAGTT
GGTATTTAAT GCCAAACAGG TGAATGCAAC CTCTCGCTCG TTACTAAGCA GGAGATAGTA
AAGTTGCTTG AAAGAGAGTT TGTTAATCAG TATAAGTAGG CTAAAGTGAG
AATATATATC TATTATTATC GGTAATGATA CTATTATTGA GAATTATTGT AGTGGGGATA
AAAATAATTT TTGGTGATTT TATCGTCCGA CTTAAAGGTG GGTTAAAAAA
GTACTTATAT TCTTTTAGAA TTGATGAAAA ATATGGGGGA ATATAATATT TATAGGAGAT
ACGATGACTA GAGTAGAGTT GATTACTAGA GAATTTTTTA AGAAGAATGA
AGCAACCAGT AAATATTTTC AGAAGATAGA ATCAAGAAGA GGTGAATTAT TTATTAAATT
CTTTATGGAT AAGTTACTTG CGCTTATCCT ATTATTGCTA TTATCCCCAG
TAATCATTAT ATTAGCTATT TGGATAAAAT TAGATAGTAA GGGGCCAATT TTTTATCGCC
AAGAACGTGT TACGAGATAT GGTCGAATTT TTAGAATATT TAAGTTTAGA
ACAATGATTT CTGATGCGGA TAAAGTCGGA AGTCTTGTCA CAGTCGGTCA AGATAATCGT
ATTACGAAAG TCGGTCACAT TATCAGAAAA TATCGGCTGG ACGAAGTGCC
CCAACTTTTT AATGTTTTAA TGGGGGATAT GAGCTTTGTA GGTGTAAGAC CAGAAGTACA
AAAATATGTA AATCAGTATA CTGATGAAAT GTTTGCGACG TTACTTTTAC
CTGCAGGAAT TACTTCACCA GCGAGTATTG CATATAAGGA TGAAGATATT GTTTTAGAAG
AATATTGTTC TCAAGGCTAT AGTCCTGATG AAGCATATGT TCAAAAAGTA
TTACCAGAAA AAATGAAGTA CAATTTGGAA TATATCAGAA ACTTTGGAAT TATTTCTGAT
TTTAAAGTAA TGATTGATAC AGTAATTAAA GTAATAAAAT AGGAGATTAA
AATGACAAAA AGACAAAATA TTCCATTTTC ACCACCAGAT ATTACCCAAG CTGAAATTGA
TGAAGTTATT GACACACTAA AATCTGGTTG GATTACAACA GGACCAAAGA
CAAAAGAGCT AGAACGTCGG CTATCAGTAT TTACAGGAAC CAATAAAACT GTGTGTTTAA
ATTCTGCTAC TGCAGGATTG GAACTAGTCT TACGAATTCT TGGTGTTGGA
CCCGGAGATG AAGTTATTGT TCCTGCTATG ACCTATACTG CCTCATGTAG TGTCATTACT
CATGTAGGAG CAACTCCTGT GATGGTTGAT ATTCAAAAAA ACAGCTTTGA
GATGGAATAT GATGCTTTGG AAAAAGCGAT TACTCCGAAA ACAAAAGTTA TCATTCCTGT
TGATCTAGCT GGTATTCCTT GTGATTATGA TAAGATTTAT ACCATCGTAG
AAAACAAACG CTCTTTGTAT GTTGCTTCTG ATAATAAATG GCAGAAACTT TTTGGGCGAG
TTATTATCCT ATCTGATAGT GCACACTCAC TAGGTGCTAG TTATAAGGGA
AAACCAGCGG GTTCCCTAGC AGATTTTACC TCATTTTCTT TCCATGCAGT TAAGAATTTT
ACAACTGCTG AAGGAGGTAG TGTGACATGG AGATCACATC CTGATTTGGA
TGACGAAGAG ATGTATAAAG AGTTTCAGAT TTACTCTCTT CATGGTCAGA CAAAGGATGC
ATTAGCTAAG ACACAATTAG GGTCATGGGA ATATGACATT GTTATTCCTG
GTTACAAGTG TAATATGACA GATATTATGG CAGGTATCGG TCTTGTGCAA TTAGAACGTT
ACCCATCTTT GTTGAATCGT CGCAGAGAAA TCATTGAGAA ATACAATGCT
GGCTTTGAGG GGACTTCGAT TAAGCCGTTG GTACACCTGA CGGAAGATAA ACAATCGTCT
ATGCACTTGT ATATCACGCA TCTACAAGGC TATACTTTAG AACAACGAAA
TGAAGTCATT CAAAAAATGG CTGAAGCAGG TATTGCGTGC AATGTTCACT ACAAACCATT
ACCTCTTCTC ACAGCCTACA AGAATCTTGG TTTTGAAATG AAAGATTTTC
CGAATGCCTA TCAGTATTTT GAAAATGAAG TTACACTGCC TCTTCATACC AACTTGAGTG
ATGAAGATGT GGAGTATGTG ATAGAAATGT TTTTAAAAAT TGTTAGTAGA
GATTAGTTAT TTTGGAAGGA GATATGGTGG AAAGAGATAT GGTGGAAAGA GACACGTTGG
TATCTATAAT AATGCCCTCG TGGAATACAG CTAAGTATAT ATCTGAATCA
ATCCAGTCAG TGTTGGACCA AACACACCAA AATTGGGAAC TTATAATCGT TGATGATTGT
TCTAATGACG AAACTGAAAA AGTTGTTTCG CATTTCAAAG ATTCAAGAAT

DNA Serotype 7

Fig. 6

AAAGTTTTTT AAAAATTCGA ATAATTTAGG GGCAGCTCTA ACACGAAATA AGGCACTAAG
AAAAGCTAGA GGTAGGTGGA TTGCGTTCTT GGATTCAGAT GATTTATGGC
ACCCGAGTAA GCTAGAAAAA CAGCTTGAAT TTATGAAAAA TAATGGATAT TCATTTACTT
ATCACAATTT TGAAAAGATT GATGAATCTA GTCAGTCTTT ACGTGTCCTG
GTGTCAGGAC CAGCAATTGT GACTAGAAAA ATGATGTACA ATTACGGCTA TCCAGGGTGT
TTGACTTTCA TGTATGATGC AGACAAAATG GGTTTAATTC AGATAAAAGA
TATAAAGAAA AATAACGATT ATGCGATATT ACTTCAATTG TGTAAGAAGT ATGACTGTTA
TCTTTTAAAT GAAAGTTTAG CTTCGTATCG AATTAGAAAA AA

Fig. 6 cont.

AAHKHVPLME YNPHEAVKNN IFGTKNVAEA AKTAKVAKFV MVSTDKAVNP PNVMGATKRV
AEMIVTGLNE PGQTQFAAVR FGNVLGSRGS VVPLFKEQIR KGGPVTVTDF
RMTRYFMTIP EASRLVIQAG HLAKGGEIFV LDMGEPVQIL ELARKVILLS GHTEEEIGIV
ESGIRPGEKL YEELLSTEER VSEQIHEKIF VGRVTNKQSD IVNSFINGLL
QKDRNELKDM LIEFAKQE

Fig. 6 cont. CPS7E

MTRVELITRE FFKKNEATSK YFQKIESRRG ELFIKFFMDK LLALILLLLL SPVIIILAIW
IKLDSKGPIF YRQERVTRYG RIFRIFKFRT MISDADKVGS LVTVGQDNRI
TKVGHIIRKY RLDEVPQLFN VLMGDMSFVG VRPEVQKYVN QYTDEMFATL LLPAGITSPA
SIAYKDEDIV LEEYCSQGYS PDEAYVQKVL PEKMKYNLEY IRNFGIISDF
KVMIDTVIKV IK

Fig. 6 cont. CPS7F

```
MTKRQNIPFS PPDITQAEID EVIDTLKSGW ITTGPKTKEL ERRLSVFTGT NKTVCLNSAT
AGLELVLRIL GVGPGDEVIV PAMTYTASCS VITHVGATPV MVDIQKNSFE
MEYDALEKAI TPKTKVIIPV DLAGIPCDYD KIYTIVENKR SLYVASDNKW QKLFGRVIIL
SDSAHSLGAS YKGKPAGSLA DFTSFSFHAV KNFTTAEGGS VTWRSHPDLD
DEEMYKEFQI YSLHGQTKDA LAKTQLGSWE YDIVIPGYKC NMTDIMAGIG LVQLERYPSL
LNRRREIIEK YNAGFEGTSI KPLVHLTEDK QSSMHLYITH LQGYTLEQRN
EVIQKMAEAG IACNVHYKPL PLLTAYKNLG FEMKDFPNAY QYFENEVTLP LHTNLSDEDV
EYVIEMFLKI VSRD
```

Fig. 6 cont. CPS7G

MVERDMVERD TLVSIIMPSW NTAKYISESI QSVLDQTHQN WELIIVDDCS NDETEKVVSH
FKDSRIKFFK NSNNLGAALT RNKALRKARG RWIAFLDSDD LWHPSKLEKQ
LEFMKNNGYS FTYHNFEKID ESSQSLRVLV SGPAIVTRKM MYNYGYPGCL TFMYDADKMG
LIQIKDIKKN NDYAILLQLC KKYDCYLLNE SLASYRIRK

Fig. 6 cont. CPS7H

```
                                                    *
Cps2J    MEKVSIIVPI FNTEKYLREC LDSIISQSYT NLEILLIDDG SSDSSTDICL EYAEQDGRIK   60
         |   ||||||  | | ||  |   ||  | |    |||| ||   | | |  |||  ||  | ||
Cps2K    MINISIIVPI YNVEQYLSKC INSIVNQTYK HIEILLVNDG STDNSEEICL AYAKKDSRIR   60

                                          *
Cps2J    LFRLPNGGVS NARNYGIKNS TANYIMFVDS DDIVDGNIVE SLYTCLKEND SDLSGGLLAT  120
          |   ||| |  ||||||       |  | ||  ||          |     |     |  |
Cps2K    YFKKENGGLS DARNYGISRA KGDYLAFIDS DDFIHSEFIQ RL_HEAIERE NAL__VAVAG  117
```

Fig. 7

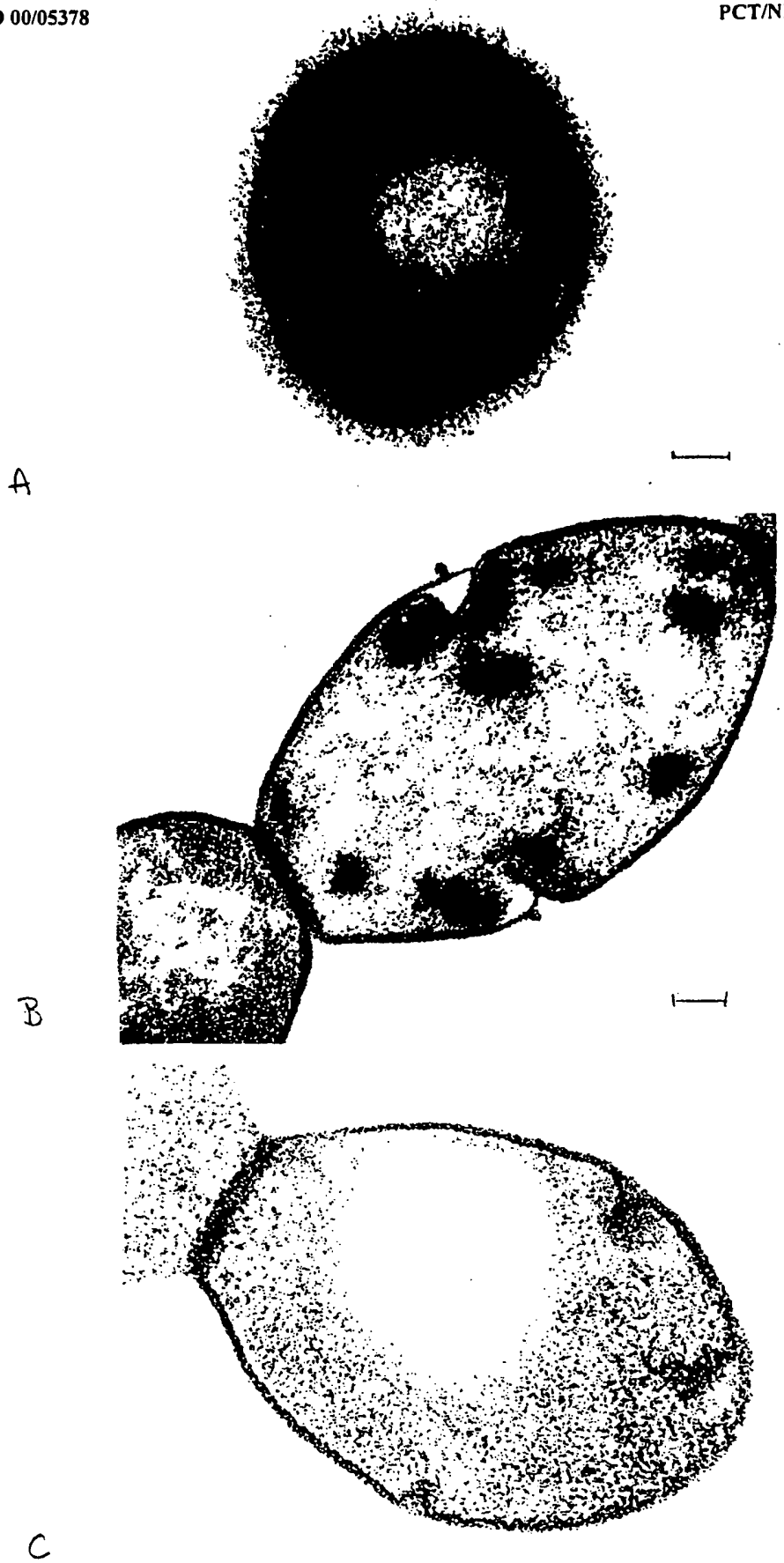

Fig. 8

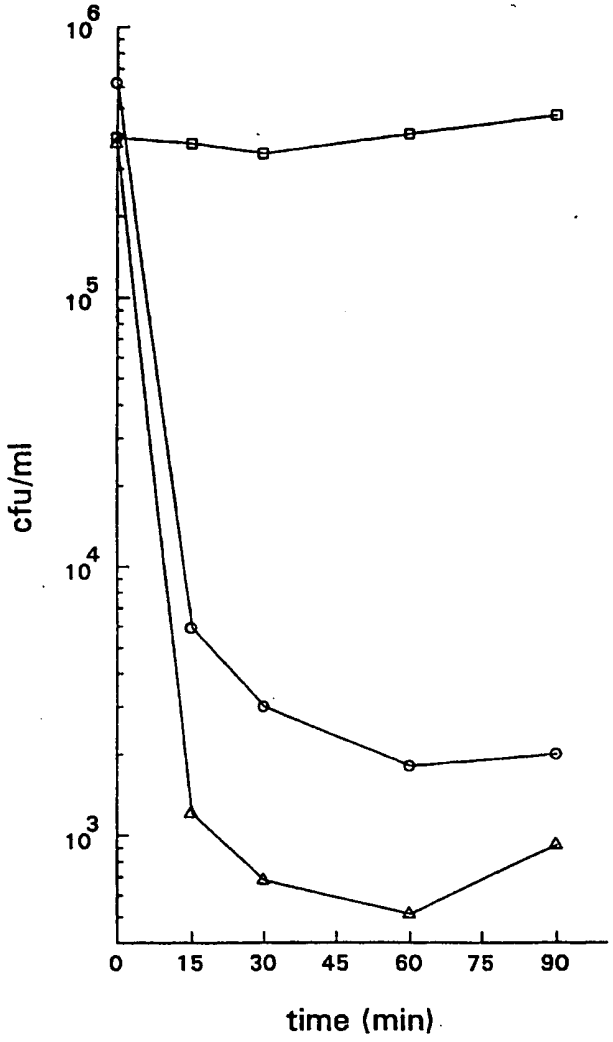

Fig. 9A

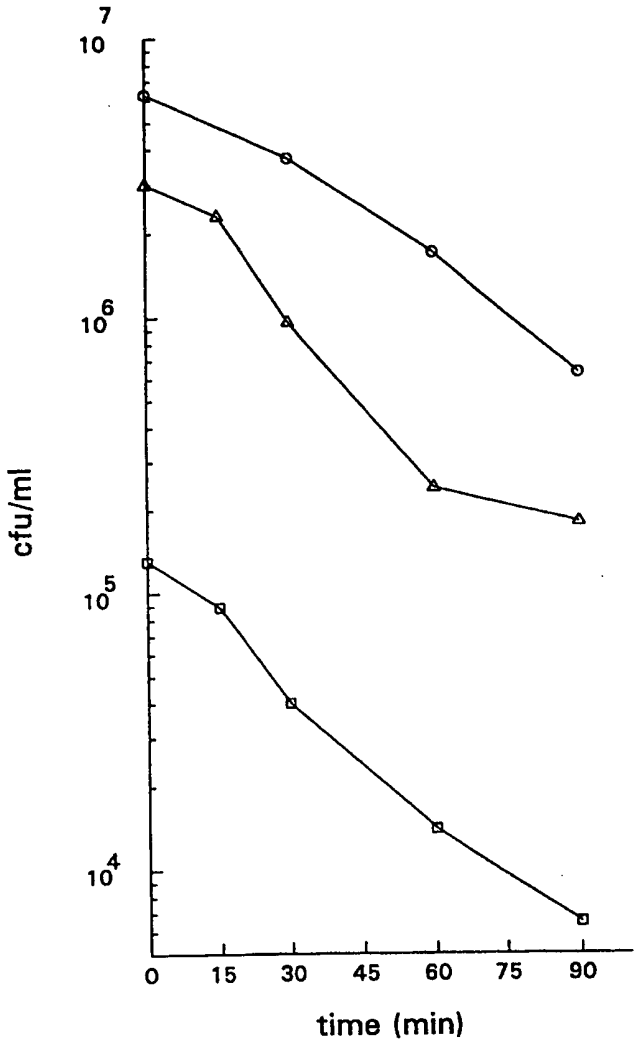

Fig. 9B

```
(1) 10508 AAGGGCACCT CTATAAACTC CCAAAATTGC GAATTTGGAG TTACGAAAGC CTTGTTAAAT CAA-CATTTA AATTTAGAA AATTAGTTTT TAGAGCTCCC  10607
          ||||||||| |||||||||| |||||||||| |||||||||| ||| ||| |||| |||||||||| |||||||||| ||||| ||||
(2) 16985 GGGGGCACCT CTATAAATTC CCAAAATTGC GAATTTGGAG TTACGAAAGC CTTGTTAAAT CAA-CATCTTA AATTTAGAA AATTAGTTTT TAGAGGTCCC  17084
          |||||||||| |||||||||| |||||||||| |||||||||| ||| ||| |||| |||||||||| |||||||||| |||||| |||||||||| ||||||||||
(3) 19803 AAGGGCACCT CTATAAACTC CCAAAATTGC GAATTTGGAG TTACGAAAGC CTTGTTAAAT CAAACATTTA AATTTAGAA AATTAGTTTT TAGAGGTCCC  19903
```

Fig. 10

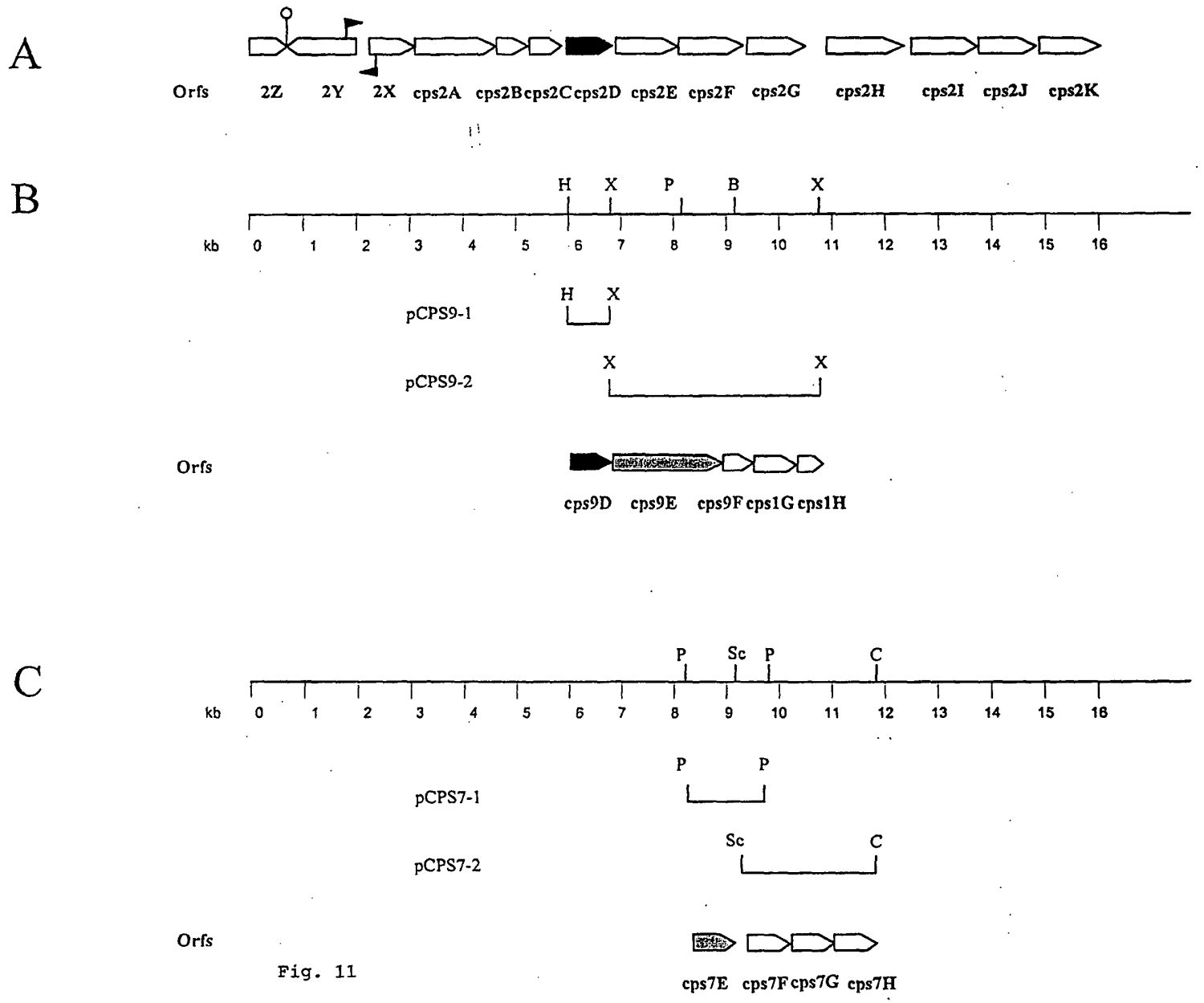
A
Orfs
2Z
2Y
2X
cps2A
cps2B
cps2C
cps2D
cps2E
cps2F
cps2G
cps2H
cps2I
cps2J
cps2K
B
H
X
P
B
X
kb
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
pCPS9-1
H
X
pCPS9-2
X
X
Orfs
cps9D
cps9E
cps9F
cps1G
cps1H
C
P
Sc
P
C
kb
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
pCPS7-1
P
P
pCPS7-2
Sc
C
Orfs
cps7E
cps7F
cps7G
cps7H

Fig. 11

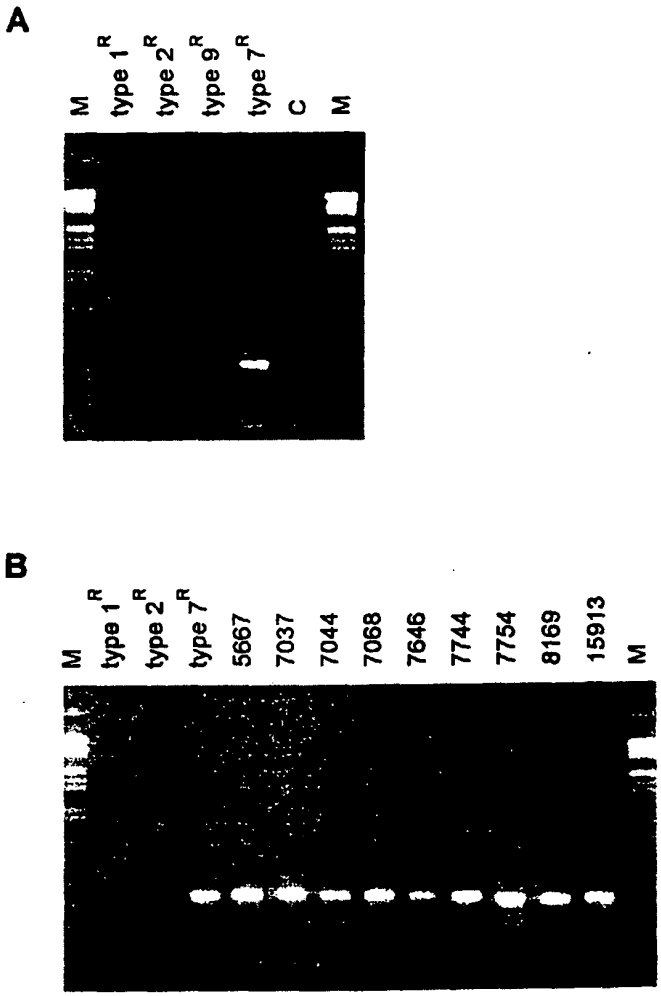
A
M
type 1R
type 2R
type 9R
type 7R
C
M
B
M
type 1R
type 2R
type 7R
5667
7037
7044
7068
7646
7744
7754
8169
15913
M

Fig. 12